

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2005, 00:57:16 ; Search time 194 Seconds
(without alignment)
3188.617 Million cell updates/sec

Title: US-09-889-325-4
Perfect score: 6424
Sequence: 1 MERLRDVRERLQAMERAFRR.....HLSPHALVGLATEELLQVAR 1208

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6424	100.0	1208	1 RCQ4 HUMAN	094761 homo sapien
2	3988	62.1	1216	2 Q75NR7	075nr7 mus musculu
3	3944	61.4	1216	2 Q99PV9	099pv9 mus musculu
4	3923	61.1	741	2 Q96F55	096f55 homo sapien
5	3469	54.0	652	2 Q96DW2	096dw2 homo sapien
6	3439	53.5	1151	2 Q76MT1	076mt1 mus musculu
7	1723.5	26.8	1579	2 Q9VSE6	09vse6 drosophila
8	1713.5	26.7	1552	2 Q7PMM8	07pmm8 anopheles g
9	1680.5	26.2	1530	2 Q9NH11	09nh11 drosophila
10	1172	18.2	340	2 Q7YR85	07yr85 bos taurus
11	971.5	15.1	927	2 Q7X829	07x829 oryza sativ
12	897.5	14.0	870	2 Q8W028	08w028 arabidopsis
13	872	13.6	941	2 Q9C6N0	09c6n0 arabidopsis
14	772	11.0	899	2 Q8ARF1	08arf1 oryza sativ
15	761	11.8	874	2 Q7FAH0	07fah0 oryza sativ
16	635	9.9	704	2 Q7MID7	07mid7 vibrio vuln
17	631.5	9.8	604	2 Q83IW3	083iw3 escherichia
18	631.5	9.8	608	1 RECO ECOGL	083iw3 escherichia
19	631.5	9.8	611	2 Q7UB49	07ub49 shigella fl
20	631.5	9.8	611	2 Q8FBM6	08fbm6 escherichia
21	631.5	9.8	611	2 Q8X8N1	08x8n1 escherichia
22	626	9.7	639	2 Q8BD16	08bd16 vibrio vuln
23	617.5	9.6	654	2 Q87MB5	087mb5 vibrio para
24	613.5	9.6	608	2 Q8CZH9	08czh9 erwinia car
25	613	9.5	642	2 Q87U96	087u96 pseudomonas
26	609.5	9.5	609	2 Q823B2	0823b2 salmonella
27	607.5	9.5	608	1 RECO SALTLY	0823b2 salmonella
28	604	9.4	644	2 Q8BEK1	08bek1 shewanella
29	599	9.3	746	2 Q7NHA8	07nha8 gloeobacter
30	593	9.2	610	2 Q6FFY3	06ffv3 yersinia ps
31	593	9.2	610	2 Q8ZAG8	08zag8 yersinia pe

32	586.5	9.1	608	2 Q7MY05	Q7my05 photorhabdu
33	586	9.1	602	2 Q88Y35	Q88y35 lactobacill
34	584.5	9.1	641	2 Q6LXK0	Q6lxk0 photobacter
35	581.5	9.1	620	2 Q9KVF0	Q9kvf0 vibrio chol
36	580.5	9.0	749	2 Q7UR05	Q7ur05 rhodospirill
37	578.5	9.0	632	1 RECO PASMU	Q9c121 pasteurella
38	578	9.0	619	1 RECO HABIN	P71359 haemophilus
39	573.5	8.9	619	1 Q6LIP9	Q6lip9 photobacter
40	573	8.9	637	2 Q6SQS2	Q6sqs2 manheimia
41	571.5	8.9	1002	2 Q7UEX6	Q7uex6 rhodospirill
42	570	8.9	603	2 Q74ER2	Q74er2 geobacter s
43	566.5	8.8	611	2 Q87KH0	Q87kh0 vibrio para
44	560.5	8.7	601	2 Q7VNM7	Q7vnm7 haemophilus
45	560.5	8.7	602	2 Q8UJ77	Q8uj77 agrobacteri

ALIGNMENTS

RESULT 1
RCQ4_HUMAN STANDARD; PRT; 1208 AA.
ID RCQ4_HUMAN
AC 094761;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE ATP-dependent DNA helicase Q4 (RecQ protein-like 4) (RecQ4).
GN Name=RCQ4; Synonyms=RECQ4;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=99097344; PubMed=9878247; DOI=10.1006/geno.1998.5595;
RA Kitao S., Ohnagi I., Ichikawa K., Goto M., Furuchi Y., Shimamoto A.;
RT "Cloning of two new human helicase genes of the RecQ family:
RT biological significance of multiple species in higher eukaryotes.";
RL Genomics 54:443-452(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20021764; PubMed=10552928; DOI=10.1006/geno.1999.5959;
RA Kitao S., Lindor N.M., Shiratori M., Furuchi Y., Shimamoto A.;
RT "Rothmund-Thomson syndrome responsible gene, RECQL4: genomic structure
RT and products.";
RL Genomics 61:268-276(1999).
CC -!- SURCELLULAR LOCATION: Nuclear (Potential).
CC -!- DISEASE: Defects in RECQL4 are a cause of Rothmund-Thomson
CC syndrome (RTS) [MIM:268400]. A disease characterized by
CC dermatological features such as atrophy, pigmentation, and
CC telangiectasia and frequently accompanied by juvenile cataract,
CC saddle nose, congenital bone defects, disturbances of hair growth,
CC and hypogonadism.
CC -!- SIMILARITY: Belongs to the helicase family. RecQ subfamily.
CC -!- DATABASE: NMB=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.inbiolegen.fr/services/chromcancer/Genes/RECQL4ID285.html".
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL, AB06532; BAA74453.1; -;
DR EMBL, AB02545; BAA6899.1; -;
DR Genew; HGNC:9949; RECQL4.
DR MIM; 603780; -;
DR WU; 268400; -;
DR GO; GO:0003678; F:DNA helicase activity; TAS.
DR GO; GO:0007275; P:development; TAS.

DR GO:0006281; P:DNA repair; TAS.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002464; DEAD box.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004589; RecQ.
 DR Pfam: PF00270; DEAD; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR TIGRFAMs: TIGR00614; recQ fam; 1.
 DR PROSITE: PS00690; DEAD ATP HELICASE; FALSE NEG.
 KM ATP-binding; Cataract; Helicase; Hydrolyase; Nuclear protein.
 FT NP_BIND 502 509 ATP (Potential).
 FT SITE 605 608 DEAD box.
 SQ SEQUENCE 1208 AA; 133076 MW; CB809A7765AB48A1 CRC64;

Query Match 100.0%; Score 6424; DB 1; Length 1208;
 Best Local Similarity 100.0%; Pred. No. 1.4e-297;
 Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MERLDVRERLQAMERAFRRORGRPSODVEAAPETRALYREYRTLRKTTGQAGGLR 60
 DB 1 MERLDVRERLQAMERAFRRORGRPSODVEAAPETRALYREYRTLRKTTGQAGGLR 60

QY 61 SSESIPAAAEAPERCWPHLNRAATKSPPTPGSRGSGVPDYGQRLKANLKTGLQNG 120
 DB 61 SSESIPAAAEAPERCWPHLNRAATKSPPTPGSRGSGVPDYGQRLKANLKTGLQNG 120

QY 121 PALGRRPWPLGRASSKASTPKPGTGPVSPFAEKVSDDEPPQLPEPQPRGRLQHLQASIS 180
 DB 121 PALGRRPWPLGRASSKASTPKPGTGPVSPFAEKVSDDEPPQLPEPQPRGRLQHLQASIS 180

QY 181 QRLGSLDPGWLQRCHEVDPFLGAPKACRPDLGSEBSQLLIPGESAVLGPAGSGQPEAS 240
 DB 181 QRLGSLDPGWLQRCHEVDPFLGAPKACRPDLGSEBSQLLIPGESAVLGPAGSGQPEAS 240

QY 241 AQOEVSIKRGSPQSSSGEKKRAMEEPESPAQVOQESSQAGPSEEGANAVVEDPPG 300
 DB 241 AQOEVSIKRGSPQSSSGEKKRAMEEPESPAQVOQESSQAGPSEEGANAVVEDPPG 300

QY 301 EPVQAQPPQCSSPSNPRYHGLSPSSQARAGKAGTAAPHITPRLARHNGYVRLNMQ 360
 DB 301 EPVQAQPPQCSSPSNPRYHGLSPSSQARAGKAGTAAPHITPRLARHNGYVRLNMQ 360

QY 361 KHYVGRALRSRLRKQAWKQKWKKEGFCGGGATVTTKESCFINEQFDHMAAQCPRA 420
 DB 361 KHYVGRALRSRLRKQAWKQKWKKEGFCGGGATVTTKESCFINEQFDHMAAQCPRA 420

QY 421 SEEDTDVGPPEPLVPSPQVPEPSLDPTVLPYLSGSPGQLAETPAEVQALEQGHQA 480
 DB 421 SEEDTDVGPPEPLVPSPQVPEPSLDPTVLPYLSGSPGQLAETPAEVQALEQGHQA 480

QY 481 FPPQGERAWRILSGISTLLVLPTGAGSLCYQLPALYSRRSPCLTLVSPPLSLMDQ 540
 DB 481 FPPQGERAWRILSGISTLLVLPTGAGSLCYQLPALYSRRSPCLTLVSPPLSLMDQ 540

QY 541 VSGLPPLCLACIHSGMTKQKQESVLOKIRAAQVHVLMTPEALVAGAGLPAPAAQLPPVA 600
 DB 541 VSGLPPLCLACIHSGMTKQKQESVLOKIRAAQVHVLMTPEALVAGAGLPAPAAQLPPVA 600

QY 601 FACIDAHCLSQMSNHPRCYLAVCVLAERMGVHFLGATATARTSDVAQHLAAVE 660
 DB 601 FACIDAHCLSQMSNHPRCYLAVCVLAERMGVHFLGATATARTSDVAQHLAAVE 660

QY 661 EBDLGPAPVPTNLHLSVSMRDPTDQALLTLQKGFQNDLSIIYCNREDTERTAALL 720
 DB 661 EBDLGPAPVPTNLHLSVSMRDPTDQALLTLQKGFQNDLSIIYCNREDTERTAALL 720

QY 721 RTCLHAAMVPGSGRAPKTTAAEYHAGMSRERRVQRAFMOQLRVVVAATVAFGGLDR 780
 DB 721 RTCLHAAMVPGSGRAPKTTAAEYHAGMSRERRVQRAFMOQLRVVVAATVAFGGLDR 780

QY 781 PDVRAVLHGLPPSPFSYVQAVRAGRDGPALCHFLPQOGDLAELARHVAADSTDFL 840
 DB 781 PDVRAVLHGLPPSPFSYVQAVRAGRDGPALCHFLPQOGDLAELARHVAADSTDFL 840

QY 841 AVKRLVQVFPACTCTCTRPSPBOGAVGGERPVKYPPOEABQLSHQAAPRFRVCMGH 900
 DB 841 AVKRLVQVFPACTCTCTRPSPBOGAVGGERPVKYPPOEABQLSHQAAPRFRVCMGH 900

QY 901 ERALPIQLTVQALMDPEAIEITLLCYLELHHHMLIELATTYTHRLNCPGSPALQALA 960
 DB 901 ERALPIQLTVQALMDPEAIEITLLCYLELHHHMLIELATTYTHRLNCPGSPALQALA 960

QY 961 HRCPLAVCLAQQLPEPDGQSSSVFPMVKLVDSMGWELASVRBALCOLQMDHPRPGV 1020
 DB 961 HRCPLAVCLAQQLPEPDGQSSSVFPMVKLVDSMGWELASVRBALCOLQMDHPRPGV 1020

QY 1021 RRGTVLVEFSELAFTLHLSPGDLTAEEKQICDPLYGAVQAEERQALRLRRTFOAFHSV 1080
 DB 1021 RRGTVLVEFSELAFTLHLSPGDLTAEEKQICDPLYGAVQAEERQALRLRRTFOAFHSV 1080

QY 1081 AFPSGCPCLQODDERSTRDLQGRYEEBEGQPGGMEDAQPEPQARLQMDWDQVR 1140
 DB 1081 AFPSGCPCLQODDERSTRDLQGRYEEBEGQPGGMEDAQPEPQARLQMDWDQVR 1140

QY 1141 CDIRQFLSLRPEKFSRAVARIFHGSPCYPAQVYGQDRFMRKYLHLSFHALVGLAT 1200
 DB 1141 CDIRQFLSLRPEKFSRAVARIFHGSPCYPAQVYGQDRFMRKYLHLSFHALVGLAT 1200

QY 1201 BELLOVAR 1208
 DB 1201 BELLOVAR 1208

RESULT 2
 075NR7
 ID 075NR7 PRELIMINARY; PRT; 1216 AA.
 AC 075NR7;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE RecQ helicase protein-like 4.
 GN Name=Recq14;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ohnata T., Fukumura R., Araki R., Abe M.;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB175741; BAF14289.1; -;
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
 DR GO; GO:0016787; F:hydrolyase activity; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0006259; P:DNA metabolism; IEA.
 DR InterPro: IPR002086; Aldehyd_dehydrog.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR011545; DEAD/DEAH_N.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004589; RecQ.
 DR InterPro: IPR001424; SOD_CU_ZN.
 DR InterPro: IPR001878; Znf_CCHC.
 DR Pfam: PF00270; DEAD; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00098; zf_CCHC; 1.
 DR PRINTS: PR00939; C2HCZFNINGER.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELIC; 1.
 DR SMART; SM00343; Znf_C2HC; 1.
 DR TIGRFAMs; TIGR00614; recQ fam; 1.
 DR PROSITE; PS00070; ALDEHYDE DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
 DR PROSITE; PS00158; ZF_CCHC; 1.
 KM ATP-binding; Helicase; Hydrolyase.
 SQ SEQUENCE 1216 AA; 135081 MW; 0A7336EBAE167306 CRC64;

Query Match 62.1%; Score 3988; DB 2; Length 1216;
 Best Local Similarity 64.4%; Pred. No. 1.1e-181;
 Matches 801; Conservative 112; Mismatches 259; Indels 72; Gaps 13;

1 MERLADVBERLOAMEARARORRRPSODDVEAAEETRALYREYRTTKRTTGAAGGIR 60
 1 MERLATVABARLOEMERARFARLRGRPAKGDVBAEETRALYREYRNKQAVROADDNR 60
 61 SSE-SLPAABEAPRRCWGHPLNPAATKSPOPTGRSRQGSVPDYGORLANKLGTLOA 119
 61 VLEQSLAAEBAEAPSCWGHPLSRAATQNTQSMPEKQSLSSVQYQKRLKANKLTQOT 120
 120 GPALGRRWPLGRASASTPKPRCTGVPVPSAEKVSDEPDPOLPFPORPRGLQHLQASL 179
 121 GPTOSRKLQLOKRSISTVPAPRPPGSKTESPCPDADDLAPVPPRPRPLGLOOLRSSL 180
 180 SQRLGSLPQWMLQCHSEVPDFLGA PKACRPDLGSEESQLIPGESAVLGPAGS----Q 235
 181 SRRLTSLDPGMERCHNRVSDLEVPAGCGLDLSAEESQPMKSKVNIADPDIOGEVSQ 240
 236 GREASAFGEVSIRVSPQSSSGEKKRWNEBPWESPAAVOQESSQAGPSSGAGAVAVE 295
 241 SPEALAQPAQVLSQSPKSKNGKRRKRNK-GEIPAQ-DQPSGAGPLSGABATVHG 298
 296 EDPPEEPVQAPORPCSSPSNRVHGLSSQABAGKAGTRPLIFPLAHDRGNVYR 355
 299 QDPPEPVOVNVQPCNS-----SNQARTKAGTTHLASPPPALDRGNVYR 347
 356 LNMKQHYVGRGALSRRLRKQAMKQMKKECEGSGGATVTTESCFPLNQPFHMAAQ 415
 348 LNMKQKRVYRVANGRLLRKQVWKQMKKQAAFGSGSPRATDQDTCRCGQFGHMASQ 407
 416 CPERA-----SEEDTAVP-----EPVSPQVPEVBSL 446
 408 CSQPEPTLVQEEGRDQDKOPISTLEBAVQRTGTASCHHSGETQPAABEQ-VPHCP- 465
 447 DPTVPLVLSLGSGLAETPAVFPQALBQAGQAPROGERAVNMLIGISTLVLPFCA 506
 466 --PMKPLVPGLGVAETPAVFPALERLGRARPCGERIMILIGISTLVLPFCA 523
 507 GKSLCYQLPALIYSRRSPCLTLVPSPLSLMDQVSGLPCLKAACIHSGMTKRORESYL 566
 524 GKSLCYQLPALIYARSPECLTLVPSPLSLMDQVSDLPCLKAACIHSGMTKRORESYL 583
 567 QKIRAAQVHMLTPEALVGA-----GGLPRAAQLPVPAFACIDEAHCLSQNSHFRPCY 621
 584 KKVRAAQVHVLIVSPEALVGCAGRPGSLPQAQLPPIAFACIDEVHCLSQNSHFRPCY 643
 622 LRVGVLEBRMGVHFLGITATATRTASDVQHLAVAEPRDLHGPAVPTLHLHSVMD 681
 644 LRVGVLEBRMGVHFLGITATATRTASDVQHLAVAEPRDLHGPAVPTLHLHSVMD 703
 682 RDTPOALTLLOGKFFONLDSIIICYNRREDTERIAALLRTCLHAAWVSGSGRAPKTTA 741
 704 RDSDALVTLLQSDFRITLDSVIITCRKQTERVAALLRTCLSNVGSRRPGCEBALA 763
 742 EAYHAGMSRERRVQORAFMOGOLRVVAVAFMGCLDRPDVRAVLTGLPPSFESYVOA 801
 764 EAYHAGMSRERRVQORAFMRGHLRVAVAFMGCLDRPDVRAVLTGLPPSFESYVOA 823
 802 VGRARDDQAPACHFLQPOGEBDLAELRHVADSDTLAVKLVQVQVPACTCTCTRRP 861
 824 IGRARDDQAPACHFLQPOGEBDLAELRHVADSDTLAVKLVQVQVPACTCTCTRRP 878
 882 SEQEGVAGGERPVKYPQEAQLS-HQAPARRVCMGHEBALPTOLTVALDMPREAI 920
 879 -----GRVSSSPEVEKESGQGYVPLGQACLGHEBALPVQSTVQQLDMPREAI 929
 921 ETLICYLELHPHMLLELATYTHCLNCPGAPQLOALAHRCPLAVCLAQOLPBDPQ 980
 930 ETLICYLELHPHMLLELATYTHCLNCPGAPQLOALAHRCPLAVCLAQOLPBDPQ 989

981 GSSVPEPMKLVDSMGWELASVRALCOLQMDHEPRTGVRGTVLVEFSELAFTLRSP 1040
 990 GRSSIEFGEVVELADSMGKLASVRALHQLKMDPEPKGAAGCTVLYKFSELAFHLNR 1049
 1041 GDLTAEERQICDPLFYGVOAERQALRLRTTQAFHSVAFPSGCPCLCEQDEERSTRL 1100
 1050 GDLTAEERQICDPLFYNVQAREHKLALHLQMSVAFPSVAFPSGCPCLCEQDEERSTR 1109
 1101 KDLGRYEEEREGEPGMEQAOGEPEQOALQMDQVDCDIQGLSLRPEKSSRAV 1160
 1110 KTVLSYEEEBE-EETETMTQGPQOTLOQMDQVDCDIQGLSLRPEKSSRAV 1168
 1161 ARIFGIGSPCYPAOVYGDPRFMRKYLHLSFHALVGLATEEL 1204
 1169 ARIFGIGSPCYPAOVYGLDRFMRKYLHLSFHALVGLATEEL 1212

RESULT 3
 ID 099PV9 PRELIMINARY; PRT, 1216 AA.
 AC 099PV9;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DS RecQ helicase protein-like 4.
 GN Name=RecQ4; Synonyms=RECQ4;
 OS Mus musculus (Mouse);
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Abe M., Ohnata T.;
 RL Submitted (May 2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB042529; BAB32696.1;
 DR HSSP; P15043; 10YY.
 DR MGD; MGI:1931028; RecQ4.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
 DR GO; GO:0016787; F:hydrolyase activity; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0006259; P:DNA metabolism; IEA.
 DR InterPro; IPR02086; Aldehyd_dehydrog.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR011545; DEAD/DEAH_N.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004589; RecQ.
 DR InterPro; IPR001424; SOD CU ZN.
 DR InterPro; IPR001878; Znf_CCHC.
 DR Pfam; PF00270; DEAD. 1.
 DR Pfam; PF00271; Helicase_C. 1.
 DR Pfam; PF00098; ZF-CCHC. 1.
 DR PRINTS; PR00939; C2HCZNFINGER.
 DR SMART; SM00487; DEXDC. 1.
 DR SMART; SM00490; HELIC. 1.
 DR SMART; SM00343; ZNF_C2HC. 1.
 DR TIGFams; TIGR00614; recQ_fam. 1.
 DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE; PS00087; SOD CU ZN. 1; UNKNOWN_1.
 DR PROSITE; PS00158; ZF_CCHC. 1.
 KW ATP-binding; Helicase; Hydrolyase.
 SQ SEQUENCE 1216 AA; 135122 MW; 431D79843BAEC2AF CRC64;

Query Match 61.4%; Score 3944; DB 2; Length 1216;
 Best Local Similarity 63.9%; Pred. No. 1.4e-179;
 Matches 795; Conservative 110; Mismatches 267; Indels 72; Gaps 13;

1 MERLADVBERLOAMEARARORRRPSODDVEAAEETRALYREYRTTKRTTGAAGGIR 60
 1 MERLATVABARLOEMERARFARLRGRPAKGDVBAEETRALYREYRNKQAVROADDNR 60


```

QY 468 EVFQALQLEQHOAFRPGQERAVNRILSGISTLLVLTPTGAGKSLCYLPALILYSRSPCLT 527
DB 1 EVFQALQLEQHOAFRPGQERAVNRILSGISTLLVLTPTGAGKSLCYLPALILYSRSPCLT 60
QY 528 LVSPILLSIMDDQVSGLPCLKAACHISGMRKQRESVLOKIRAAQVHVLMTPEALVGA 587
DB 61 LVSPILLSIMDDQVSGLPCLKAACHISGMRKQRESVLOKIRAAQVHVLMTPEALVGA 120
QY 588 GGLPPRAQOLPVAFAACIDBAHCLSQSMNRPCTYLAVCVLAERMGVHFLGTATATXR 647
DB 121 GGLPPRAQOLPVAFAACIDBAHCLSQSMNRPCTYLAVCVLAERMGVHFLGTATATXR 180
QY 648 TASDVAQHILVAEPEPLHGPAPVPTLHLSVSMNDRTDQALLTLQKGFONLDSITTC 707
DB 181 TASDVAQHILVAEPEPLHGPAPVPTLHLSVSMNDRTDQALLTLQKGFONLDSITTC 240
QY 708 NRREDTERIALLRCTLHAAMVPGSGRAPKTTAEVYHAGMSRERRVQRAFMQGLAV 767
DB 241 NRREDTERIALLRCTLHAAMVPGSGRAPKTTAEVYHAGMSRERRVQRAFMQGLAV 300
QY 768 VVATVAFGGLDRPDVRAVLHLGLPSPFESYVQAVRAGRDGPANCHLFLQPGEDLRE 827
DB 301 VVATVAFGGLDRPDVRAVLHLGLPSPFESYVQAVRAGRDGPANCHLFLQPGEDLRE 360
QY 828 LRRHVADSTDFLAVKRLVQVVPACTCTPRPSQEGAVGGERPVKYPPOEALQSH 887
DB 361 LRRHVADSTDFLAVKRLVQVVPACTCTPRPSQEGAVGGERPVKYPPOEALQSH 420
QY 888 QAPRRVRCMGHERALPIQLTVQALDMPEALITLVCYLELPHHMLTATYTHCHL 947
DB 421 QAPRRVRCMGHERALPIQLTVQALDMPEALITLVCYLELPHHMLTATYTHCHL 480
QY 948 NCPGPAQLOALAHRCPLAVCLAOQLPEDPGQSSSVDPVYKVLVDSMGWELASVRL 1007
DB 481 NCPGPAQLOALAHRCPLAVCLAOQLPEDPGQSSSVDPVYKVLVDSMGWELASVRL 540
QY 1008 COLQMDHEPRTGVRGTGLVFEFSLAFHLRSKGLDTAEKQICDPLGROYARRQAL 1067
DB 541 COLQMDHEPRTGVRGTGLVFEFSLAFHLRSKGLDTAEKQICDPLGROYARRQAL 600
QY 1068 ARLRRTFQAFHSVAFSPGCPCLQODEERSTRKLDLGRVFEFEEBQEPGEMDAGPSP 1127
DB 601 ARLRRTFQAFHSVAFSPGCPCLQODEERSTRKLDLGRVFEFEEBQEPGEMDAGPSP 660
QY 1128 GQARLQDMEDQVRCIDRQFLSLRPEEKFSRAVARIFHGISPCYQAVYGGDRFRWKY 1187
DB 661 GQARLQDMEDQVRCIDRQFLSLRPEEKFSRAVARIFHGISPCYQAVYGGDRFRWKY 720
QY 1188 LHLSPHALVGLATTEELQVAVR 1208
DB 721 LHLSPHALVGLATTEELQVAVR 741

```

```

RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stranstein M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Cantucci P., Prange C.,
RA Rana S.B., Loguettano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.V., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalios D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.,
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.",
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013277; AAH13277.2; -.
DR HSSP; P15043; 10YY.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0008026; P:ATP-dependent helicase activity; IEA.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00490; HELIC_C; 1.
SQ SEQUENCE 652 AA; 72987 MW; 328B7B4BD39735F7 CRC64;

Query Match 54.0%; Score 3469; DB 2; Length 652;
Best Local Similarity 100.0%; Pred. No. 2.8e-157;
Matches 652; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 MTRKRESVLOKIRAAQVHVLMTPEALVGAAGLPPAOLPVAACIDEAHCLSQSMN 616
DB 1 MTRKRESVLOKIRAAQVHVLMTPEALVGAAGLPPAOLPVAACIDEAHCLSQSMN 60
QY 617 FRPCTLRVCKVLRERMGVHCFGLTATATRTASDVAAQHLVAEPEPLHGPAPVPTNL 676
DB 61 FRPCTLRVCKVLRERMGVHCFGLTATATRTASDVAAQHLVAEPEPLHGPAPVPTNL 120
QY 677 SVSMNDRTDQALLTLQKGFONLDSITTCNRREDTERIALLRCTLHAAMVPGSGRA 736
DB 121 SVSMNDRTDQALLTLQKGFONLDSITTCNRREDTERIALLRCTLHAAMVPGSGRA 180
QY 737 PRTAEAYHAGMSRERRVQRAFMQGLRVVAVVAVAGMGLDRPDVRAVLHLGLPSP 796
DB 181 PRTAEAYHAGMSRERRVQRAFMQGLRVVAVVAVAGMGLDRPDVRAVLHLGLPSP 240
QY 797 SYVQAVRAGRDGPANCHLFLQPGEDLRELRHVHADSTDFLAVKLVQGVFPACTCT 856
DB 241 SYVQAVRAGRDGPANCHLFLQPGEDLRELRHVHADSTDFLAVKLVQGVFPACTCT 300
QY 857 CTRPSEBQGAVGSRPVPKYPPQAEOLSHQAPGPRVRCMGHERALPIQLTVQALDMP 916
DB 301 CTRPSEBQGAVGSRPVPKYPPQAEOLSHQAPGPRVRCMGHERALPIQLTVQALDMP 360
QY 917 EEAITLVCYLELPHHMLTATYTHCRNLCPGSPQLOALARCPLANCLAOQPE 976
DB 361 EEAITLVCYLELPHHMLTATYTHCRNLCPGSPQLOALARCPLANCLAOQPE 420
QY 977 DPGQSSSVFEDMVTLVDSMGWELASVRRALCOLQMDHEPRTGVRGTGLVFEFSLAF 1036
DB 421 DPGQSSSVFEDMVTLVDSMGWELASVRRALCOLQMDHEPRTGVRGTGLVFEFSLAF 480
QY 1037 LRSPEDLTAEKQICDPLVGRVQAREQALRLRRTQAFHSVAFPSGCPCLQODEER 1096
DB 481 LRSPEDLTAEKQICDPLVGRVQAREQALRLRRTQAFHSVAFPSGCPCLQODEER 540
QY 1097 STRLMDLGRVFEFEEBQEPGEMDAGPPEQALQDMEDQVRCIDRQFLSLRBEKKS 1156

```

```
Db 541 STRLMDLGRYFEEBEGQEPGGMEDAQGPBGQARLQDWEQVRCIRPQFLSLRPEKFS 600
Qy 1157 SAAVARIFHGIGSPCPYPAQVQGDRRFMRKYLHLSFHALVGLATEBLLOVAR 1208
Db 601 SAAVARIFHGIGSPCPYPAQVQGDRRFMRKYLHLSFHALVGLATEBLLOVAR 652

RESULT 6
076MT1 PRELIMINARY; PRT; 1151 AA.
ID 076MT1 AC 076MT1
AC 076MT1; 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DE RecQ4 homolog (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Abe M., Ohnata T.;
RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB039882; BAD1131.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0006026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006259; F:DNA metabolism; IEA.
DR InterPro; IPR002086; Aldehyd_dehydroy.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004589; RecQ.
DR InterPro; IPR001424; SOD_CU_ZN.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PR00098; zf-CCHC; 1.
DR PRINTS; PR00939; C2HCNFINGER.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR SMART; SM00343; Znf_C2HC; 1.
DR TIGRFAMs; TIGR00614; recQ_fam; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYD_CYS; UNKNOWN_1.
DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
DR PROSITE; PS00158; ZF_CCHC; 1.
DR ATP-binding; Helicase; Hydrolyase.
FT NON TER 1
FT NON TER 1151
SQ SEQUENCE 1151 AA; 126528 MW; 97CF481ABCA4B9A CRC64;

Query Match 53.5%; Score 3439; DB 2; Length 1151;
Best Local Similarity 59.9%; Pred. No. 1.4e-155;
Matches 707; Conservative 103; Mismatches 248; Indels 122; Gaps 14;

Qy 16 RAFFRRGRRRPSQDDVEAARPEETRALYREYRTKRTTGAAGGLRSE-SLPAARAEAR 74
Db 1 RAFFRRGRRRPAQDVEAARPEETRALYREYRTKRTTGAAGGLRSE-SLPAARAEAR 60
Qy 75 PRCWGHILNRAATKSPQPTTGRSGRQGVDPYQRLKANTLKGTLQAGPALGRRPWPLGRAS 134
Db 61 PCWGHILNRAATKSPQPTTGRSGRQGVDPYQRLKANTLKGTLQAGPALGRRPWPLGRAS 120
Qy 135 SKASTPKPPGCTGVPFSAEKVSDPEPQLPEPQPRGLQHLQASISQRLGSLDPGWLQRC 194
Db 121 STVPARPPGSKTESPCPDADALPRVPEPRRLQOLQOLSSSLRRLTSLDPGWLERC 180
Qy 195 HSEVPPLFAPKACRCDLGSSESQLLIPESAVLGGAAS----GPEASAPQGVSIIRG 250
Db 181 HNRVSLLEVPAGCGLDLSABESQPMGSKVINIADPDIOSEVSVSGPEALIAQPAQVLSQ 240
```

```
Qy 251 SPQSSSGGKRRKRNNEBPWESPAQVOQESSQAQPPSEAGAVAAVEDEPGEFVQAQPP 310
Db 241 SPKSIINSGRKRKNNEK-GEPPAQ-DQSSGAGFLSEBARATVHGQDPGEFTQVNVDP 298
Qy 311 CSSPSPNPRKHLSPESSQARAGAEETAPLHLPRLARDKRGVYVLMNKKHYVGRALR 370
Db 299 CNS-----SNQARTEKAKGTTHLHSPRASLDKRGVYIRLNNKKNRFPVAVGNR 347
Qy 371 SRLRKQAKQKWRKKKCGCGGATVTTTSCFPLNEQFDHMAACPRPA-----SEED 424
Db 348 GRLLRKQVWKQKKKKQKQAFAFGSGPRATDKDTCFRCGQFGHNASQSGPGLTVQEBGD 407
Qy 425 TDVAVP-----EPLVSPQVPEVPSLDPTVPLVLSIGPSGQ 461
Db 408 RDKQPISTLBEVAQRTGASCHSGSEBTQPAAPELQ-VPHCPF---PMSPLYPGPGLGQ 463
Qy 462 LAETPAEYFQALFQGHQAFRRGQRAWMRTISGISTLLVPTGAGKSLCYQLPALVSR 521
Db 464 VAETPAEYFQALERGYRFRGQERAIWRLISGISTLLVPTGAGKSLCYQLPALVSR 523
Qy 522 RSPCLTLVSPPLSLMDQVSGLPCLKAACIHSGMTKQRESVLOKIRAAQVHVMLTLP 581
Db 524 RSPCLTLVSPPLSLMDQVSDLPCLKAACIHSGMTKQRESVLOKIRAAQVHVMLTISP 583
Qy 582 BALVGA-----GGLPRAQLPEVAFACIDEAHCLSQWSHNRFPCLRYCKVLRBMGVHC 636
Db 584 BALVGCAGRPGPSLPQAQLPPIAFACIDEVYCLSQWSHNRFPCLRYCKVLRBMGVHC 643
Qy 637 PLGLTNTTRKASVVAQHLAAVEERDLHGPRVPTNHLSSMRDQDLTLTGGR 696
Db 644 PLGLTNTTRKASVVAQHLAAVEERDLHGPRVPTNHLSSMRDQDLTLTGGR 703
Qy 697 FQNLDSIIIVCNRRDETRIAALRTCLAAWVPSGGRAPRTAEAVHAGSCSRRRRV 756
Db 704 FRTLDVSIITCTREKIIQNGIMIALRTCLSMVGDSRPGCGPALIAEHAHAGSSGERRV 763
Qy 757 QRAFWQGLRVVAVATVAFQMGIDRPDVAVLHGLSPSPESYVQAVGRDQDPALCHL 816
Db 764 QOAFMRGHLRMVAVATVAFQMGIDRPDVAVLHGLSPSPESYVQAVGRDQDPALCHL 823
Qy 817 PLQP-----GGEIDR 826
Db 824 FWHQVGSPISPDQDRPGSTIPRLQPOLSLCLVSCRPGKCGSSVHMTVPMQGEDLM 883
Qy 827 ELRRVHADSTDFLAVKRLVQRFPACTCTTRPSEBOGAVGGRPVYKYPQOEAOLIS 886
Db 884 ELRRVHADSTDFLAVKRLVQRFPACTCTTRPSEBOGAVGGRPVYKYPQOEAOLIS 929
Qy 887 -HQAAPGRPRVCMGERALPIQLTVQALDMEBEALITLLCYL-ELPHNMLELATTYTH 944
Db 930 GQQTVPVLGQACLSGERALPVQSTVQALDMEBEATETLLCYLKLPHHMLELTPVTAQ 989
Qy 945 CRLNCPGRPAQIALAHNCPRLAVCLAQQLPEPDQSGSSVFPDVKLVDSMGWELASVR 1004
Db 990 CHLHCLGSAQIQALAHNCPRLAVCLAQQLPEPDQSGSSVFPDVKLVDSMGWELASVR 1049
Qy 1005 RALCOLMDHBEPTGVRRGTGLVFESELAFLHRSGLTLAEKQOICDPLVGRVQABER 1064
Db 1050 QALHQLKMDPEPKKAAGCTGLVYKFSLAFLHLSRGDLTDEKQOICDPLVLRQAHEN 1109
Qy 1065 QALARLRRTFOAFHSVAFPSGCPCLQODEERSTRLKDL 1104
Db 1110 KALAHLQWSKAFRSVAFPSGCPCLQESNEEHSNQKLV 1149

RESULT 7
Q9VSE6 PRELIMINARY; PRT; 1579 AA.
AC Q9VSE6;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-MAR-2004 (Tremblrel. 26, Last annotation update)
CG7487-PA.
```

GN Name=RecQ4; ORFNames=CG7487;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.
 NCBI_TaxId=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abail J.F., Adyanti A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadien E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,
 RA Hoscin D., Houston K.A., Howland T.J., Wei M.H., Idegam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Jimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshirei A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shie B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,
 RA Spletter E., Spreidling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskae R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodgett W., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Zeng R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of Drosophila melanogaster.";
 Science 287:2185-2195 (2000).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodeggen E.J.,
 RA Svirskae R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,
 RA "Finishing a whole-genome shotgun: Release 3 of the Drosophila
 melanogaster euchromatic genome sequence.";
 Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirskae R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.,
 RA "The transposable elements of the Drosophila melanogaster euchromatin:
 a genomic perspective.";
 Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,

RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale P.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.,
 RA "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review.";
 RL genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirskae R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.,
 RA "The transposable elements of the Drosophila melanogaster euchromatin:
 a genomic perspective.";
 Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 [6]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,

Query Match 26.8%; Score 1723.5; DB 2; Length 1579;
 Best Local Similarity 28.9%; Pred. No. 8-2e-74;
 Matches 467; Conservative 217; Mismatches 455; Indels 471; Gaps 45;

10 RLQAWERAAPRQGRPSQDDVEAAPEETRALYREYRTLKRT----- 51
 16 RYKWEKDFKKKNGRVPKDYRIDASGEIRDSYKMYLKTSPLEETLNDVLSDEGYDIL 75
 52 -TGQAGG-----GLRS---ESLP----- 66
 76 EMSQASDGVSMVLDQVSLNGLPGLDLSALVGQSSGNLEETPQSVGSEFSLNLDLPN 135
 67 -----AAAEAP-PPRCGPHLNAATKSPOP-----TPGRSRG 100
 136 RQVLTNLVNRDENHYRKEFAVEBELPIQNMAGLNVSKPPAPPPVVASASAPBHGQ- 194
 101 SVPDYGRLKANLKTGL-QAQPALGR--PWPGLRASSKASTPKPGTGPVSPFAKXVS 157
 195 --PKAGASLKPSSLAKLFGSSRGFAKRNPRPLSRVSSSSSTTSLSSVPTDHHBELDIF 252
 158 EPPQPEPPQPPRGRQ-----HLQASLSGRSLGSLDPCWQQR-----CHSEVP- 199
 253 ETLIRKAKQEKYKQQAIALNPMLASESKSISIKT--LVDDGWRRTKENTLDEEVPF 309
 200 ----DFLQAPK-----ACRPDLGSEESQLIPGSAVY-----GP 230
 310 AEAANNSTGSKTNRGLANLDSLKTKPTVKEKXVLAQAPDGMALIQELQTMNSMCKP 369
 231 GAGGQGPASAPQEVSVIRVSGPPQPS-----SGEKRRKRNNE----- 267
 370 DHLNHTTPASSOKSVAPK-NKPPSEGETDSDSVVASESEDEPPQRYOLSKRRKIVS 428

QY	268	-----WPESAOUQOOS-----SOAGP-----	284
Db	429	TASGKVEVAAPVEIIPNKVPEPETETPAQENPDPSADEODATYVBNKKKDKAKRKOAAKG	488
QY	285	-----PSEGAVAVAEEDPGEPPVOAOPQP----	310
Db	489	QKTKPKAPEPKTEKKAQVAEKKPKAEKKPRNSKALAVEPADPDEDEQPLNPFBDL	548
QY	311	-----CSBSPNRH-----GLSPSSQARAGKAGTAP	338
Db	549	KYVIALBAGDITTSVPRINVOOLEBADATAQRYIRTPAAGPNPGISLEGGINIRDEKRAAR	608
QY	339	LHIFRLRLRD-RGVYVRLNKKOKHYRG-RALRSRLRLKQAMKQMR-----KG	387
Db	609	KLBERIAGKLNENFTIINIOKKFVGVGKSVNFSKTKKOOMRHKKAVALISGPDMDMG	668
QY	388	ECFGGGAGVTYTKESCFLNEQFDHMAOC-----PRP-----	419
Db	669	GODG-----VLT-----CFQCGGVGHQAQCKVKGSDLSLPSAQLEBDEBPFTPLAAGEMA	721
QY	420	-----ASEBDTAVGPEPLVPSPQP-----	439
Db	722	SQGAVVAHSRNISRLPQANAALIQDELENEEEDQSSGDEVEQHPDPWMSDEMDV	781
QY	440	-----VPE-----VPSLDP-----VL	451
Db	782	DPEALDAVAESLSOPVQOEKASPIKTYVGHKIPPEPLKQAGLDTTASSNSNRSHGVK	841
QY	452	PLYSIGPESQALETPAEVAEQALEQHOAFRPGQERAVMRILSGISTLVLPYTGKSLC	511
Db	842	PLYDLTPGOSVODTTPVLEBALHMGHNFRKGQORALIMRLTSGSLSVLTSGSKSLC	901
QY	512	YQPLALYSRNSPCLTLVVSPLSLMDQVSGLPBCLAACTHSQMTKORESVLOKIRA	571
Db	902	YQPLAYLSRKVGAITLVISPLVSLMEDQVGVPHFLBAHCLHTNQTPAQRMKIQMTAN	961
QY	572	AQVHVLMTPELV-----GAGGLPRAQOLPRVAFACIDEAHCLCSQNSHNRPCYLAVC	625
Db	962	GEIDLTVSPBEAVVAGERATGFGAI--LRQPLPIAFACIDEAHCVCSQNSHNRFPSTYMIC	1019
QY	626	KVLRRMGVHCPGLGATATRTRTASDVAVQHLAAVE--EEDLGPAPVPPNHLVSQMDRT	684
Db	1020	KVLKRNIGVRYTLGLTATATLTPTRVSIINHGLISGEGKIIISDTPLPDNLVLVSQDENR	1079
QY	685	DOALLTLLOQRFPQMLDSIIIVCNREDTERIALLRKLCILAAWVPGSGGAPKT-----T	740
Db	1080	DALLQILNSERFEPQSGIIITCRDEGERIAGFIRTCVQORRREPQDTQTKRKRVAMQ	1139
QY	741	AAAHYAGMCSRERRVQORAFQOGQLRVVVATYAFKGLDRPVRVAVHLGLPSPSESTYQ	800
Db	1140	AAHYAGMCPASRRRTYOKAFMSNELRIVATIPMGGINKPDIDRIVIHYNMERPNSESTYQ	1199
QY	801	AVGRAGRGORAHCHLPLOPOGEDRLRLRBVHADSTQFLAVKRLVQVFPACTCTCPR	860
Db	1200	EIGRRGRGRLPSHCHLPFDACKGDSSELRHVYNSIDRHYIRKLQOKI FVPCSC-----	1255
QY	861	PSEOGAVGERPVKEYPQOEAQOLSHQAAQPR-RVQMGHERALPIQITVQALMDPEEA	919
Db	1255	--DKRASKRTALPIP-----LEGSGPRVHMCRGHEIGTSVEKTEVMMDIAPEN	1300
QY	920	ITFLLCYELARHMLELIATTYTCRLNCPGEGPQALQALAHRCPLPACVLAQOLPEDP-	978
Db	1301	ISTLLCYMELDPKMCISVLSASVYAKVISYGGPKYLLKHAKECPPLAMAIQLQIRDTFE	1360
QY	979	GQSSSVSEFDMVKLVDSMGWELASVRLALCOLQOMHERITGVYRKRTGVLVSESELAFLHR	1038
Db	1361	KEDSNITIESVTDIAGIGMNSGVKYOQKOLEM--VKVNGVPKSPITVSYDIDGFRIK	1418
QY	1039	SPGDLTAAEKDQICDFLGRVQARERQALARLRTFQAHSVAFSPSCPECLEQO--DEERS	1097
Db	1419	VFGDTEBESTIDNALTTLYTRSKQERTQOLIQUOYAHGLAAVAVYSSCGQCSNADFPQDRG	1476
QY	1098	TRLKULGRYFEEBEGQ-----BEGGMEDAQGPBGQARLQDMEQVRCIDIRQELSLRP	1151

```

Db      1479  EOLKAIVANYANDYPODLELEIEBSNPD-----ENTIDVHALINNP 1522
Qy      1152  EKKFSRRAVANIPIFGIGSPCTPAOYVGGDRFFRKYLHLSHALVGLATBELQ 1205
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1524  DNTFTGRNARIPIFGI MSPYPAVITWGR-CRKMRAHVVDENRILHLNMAI1K 1576

RESULT 8
Q7PNM8  PRELIMINARY; PRT, 1552 AA.
ID      Q7PNM8
AC      Q7PNM8;
DT      01-MAR-2004 (TReMBLrel. 26, Created)
DT      01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DT      01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE      ENSANGP0000004546 (Fragment).
      Name=ENSANG00000003572;
OS      Anopheles gambiae str. PEST.
OC      Eukaryota, Metazoa, Arthropoda, Hexapoda, Insecta, Pterygota,
OC      Neoptera, Endopterygota, Diptera, Nematocera, Culicoidae, Anopheles.
NCBI_TextID=180454;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=PEST;
RL      Anopheles Genome Sequencing Consortium;
RA      Submitted (Apr-2003) to the EMBL/Genbank/DBJ databases.
CC      -!- CAUTION: The sequence shown here is derived from an
      EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
      preliminary data.
CC      EMBL; AAAB01008960; EAA11913.2; -.
DR      HSSP; P15043; 10Yy.
DR      GO; GO:0005524; F:ATP binding; IEA.
DR      GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR      GO; GO:0003676; F:nucleic acid binding; IEA.
DR      GO; GO:0006259; F:DNA metabolism; IEA.
DR      InterPro; IPR01410; DEAD.
DR      InterPro; IPR011545; DEAD/DEAH_N.
DR      InterPro; IPR01650; Helicase_C.
DR      InterPro; IPR004589; RecQ.
DR      InterPro; IPR001878; Znf_CCHC.
DR      Pfam; PF00270; DEAD. 1.
DR      Pfam; PF00271; Helicase_C. 1.
DR      Pfam; PF00098; zf_CCHC. 1.
DR      PRINTS; PR00939; C2HCZNFINGR.
DR      TIGRFAMs; TIGR00614; recQ_fam. 1.
DR      PROSITE; PS0158; ZF_CCHC. 1.
DR      NON_TER 1
FT      SEQUENCE 1552 AA; 171331 MW; 35DDA746F98A1089 CRC64;

Query Match 26.7%; Score 1713.5; DB 2; Length 1552;
      Beat Local Similarity 30.0%; Pred. No. 2.4e-73;
      Matches 485; Conservative 194; Mismatches 436; Indels 503; Gaps 49;

Qy      10  RLQAWERAFRORGRSPQDVEAAPBETRLAYREYRLTKR----- 50
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      15  RVKQWEKEFRKKYGRIPSKDIRBSPSEIRNSYKLYQTKNLLKHTLLLEALDDDDPDG 74
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      51  -----TTGACGGGGLRSSSSL-----PAASEAPRPR--- 76
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      75  SLISQEASFQSDVSDVSGLLISSTINDSALLERSSMSGLTKPATAQPSPTTHEA 134
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      77  -----CWGPHLNRAATKSPQPT-----PGRSHQGVDPYQRLKANLKG--TLQAGP 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      135  EKLQDHNLVVAMQGEKSRKAVQSTKPAEITPPK-----PVLTKGITTLLPKST 181
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      122  ALGRRPWP-----LGRASSKASTPKPPTGP-----VPSFAEKYS 156
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      182  ALIPKRNKRKISIRNSFGPSSQGSST--NGTAPGTVLPDLETLITAKSKELIESNATYKE 238
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      157  DEPPDLPQRPRGTLQHLQASISGRGLSDPGCMQRC-----HSEVDDPLGAPACRP- 210
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      239  EKLPRP-----LVPTQSPINQDLBRWYNRLRISGGTGDSDVTTTGARPSSEBPT 287

```

```

QY 211 -DLGSEESOLLIPESAVLGPAGSQ-----GPEASAFQEVISIRVG----- 250
Db 288 PTNBLRNVNVAASLASATSPGMSALSTMDKAGTAVBQATA---AARIGVQNTAYDA 344
QY 251 -----SPOPSSSG---GKKRMNEBPWESPANVOQESSQAGPSEBAG 290
Db 345 AIHSDSDGVENSEDESATKTSFGMLHIAKKRKIFTSINRSPVE-RVEAKQKPSIETNN 403
QY 291 AVAVE-----EDPG---EPVQAPR-----QPC-----SSPN 316
Db 404 VVSIVSKVENODPCGVABEABQATPKGAPKRVQKRVSKAKQPKVKRSTTRKSTTIS 463
QY 317 PRYHGLSPSSQARA-----GKAEGTAP-----LHIF 342
Db 464 PK---LSFGSSKPAQGRGKAKKIKTKRKVDVGESEBAAPBAEQOYVEPGLARLSI 520
QY 343 PRLAHD-----RG 351
Db 521 PRLDINELVRDTKLADSFILGAGSABEGSSSGKTRTTVAGSTKETALRKMAAGKLINE 580
QY 352 NYVRLAMKQKHVRC-RALRSRLAKQAKQKRRKKEFGG-----GATVTTKES 402
Db 581 NFVRIDIRKRVFVKCKTINYSRYKKSQWK---AKKAALTGPBMDRCDDGTLT--- 633
QY 403 CFLNDFQFMAQCPRAPESE---DTDAVGE----- 431
Db 634 CFQCGGTGHMAQCKKPLEDKLLPYDADTIESSSTPTLEBAMANSRTLAVHSRIERL 693
QY 432 PLVPSR-----QPVPEVPSLDP----- 448
Db 694 PVNAVPMWKEQEDLPAEBEVEGEPKEPSTLPVEDLDDEPMQESDEHOKETRONAA 753
QY 449 -----TVLPYISGPGSQALETAEVFOALE 474
Db 754 SOPAPAYIGHKIPBEFLKQSGLEVTWVGSGFKGVEPLPYQORGCOTYPTPEVEFALR 813
QY 475 QLGHOAFRPGORAVRILSGISTLLVLPYTGAGKSLCYQPLALYXSRSPCLTLVSPLL 534
Db 814 MFGHOSFRRHGBRANRVLGSLTVLTSTGKSLCYQPLALYXSRSPCLTLVSPLL 873
QY 535 SLMDQVSGLPCLKAACISGQTRKQKESVLOKIRAAQVHVMJLTPALV-----GAG 588
Db 874 SLMEQVHGMPLFNAHCHTNOTKVRNRTMOALAGVDVLLISPEAVSGEKSTGCG 933
QY 569 GLPRAQLEPVAFACIDEAHCLISQNSHNPFCYLAVCYKLRMGVHCHGLTATAT-R 646
Db 934 SL--LROLPPIAFACIDEAHCLISQNSHNPFCYLAVCYKLRMGVHCHGLTATAT 991
QY 647 RTASDVAGHLAAVEBPD---LHGRAPVPTNLHLSVMDRDTQALLTLQGRFQNDLS 702
Db 992 ROAGSIVSHLAI--PDGVQGIISDIPLDNLLTLTVSRDANDVALVELLOSERSSLOS 1048
QY 703 IIVYCNREDETERIALLTCTCLHAAMVPGSGRAPPT---TAEAYHAGCSRRRRVQ 757
Db 1049 IIVYCTRDDDCRVAFTIRCTQDABABAAAANAANKRKLAVABEYHAGMASRRRTQ 1108
QY 758 RAEMQGLRVVAVATYAFGKGLDRPVRAVLAHLGLPSPFSYQVAVGRAGDCQAPACHLE 817
Db 1109 NAFMGEIRIVAVATTAFGKINKADIRAIHYNMPNFESYQVAVGRAGDGLTCHCHLF 1168
QY 818 LQPOBEDLRELBRRHNAOSTDLAVKRLVQVFPACTCTCTPRPSQ-----EGAVGGR 872
Db 1169 LDGKGTDRNELRRFTYANSIDHVIKRLQKIFVPACAC---KIQOVILLFTDGGGGGGR 1224
QY 873 PVPRKVPQBAQLSHQAAAPRRVCGHERALPIQTLVQALDMPBEAIFTLLEYLH 932
Db 1225 -----KRLCGHICISIAETVQGLDIPRENNITPLCYLELDQ 1263
QY 933 HMLELLATTYTHCRINCPCGPAQOLALAHRCPLAVCLAQOIPEPDGCG-SSSVFDMVK 991
Db 1264 RYIQALSPAYTWCKVMSYGVVRPLROAKACECPPLAMAFALDLKRGISHATSTAIERPID 1323
QY 992 LVDSMGWELASVRLALCOLQMDHBERTVGRGTGVAVERSBELAFHRSGLDLAEKDOI 1051

```

```

Db 1324 VASATGDSGVVKKQOLKLEW---TTVNVNRKRSPLSVFTFELGFRVRAPCDLTBEILDHA 1381
QY 1052 CDPLFGRVQAEERQALARTRTFOAFHSAVAPSCPCLEQOD--BEESTRLKDLGRYFE 1109
Db 1382 LDGLYERTYHBERQALQLOLYISDLNSVCENTIGP-VSRADCPPTGPKDKTYIREYFT 1440
QY 1110 EBEQEPGCMGDAQEPPEQARLDW-----EDQVRCDIROPLS 1148
Db 1441 TDISKE-----QIEIVPEVISIMIKRFXKLAVTVIPLPIQPDTTBEQLINDIRTAIC 1493
QY 1149 LRPEKFSRAVARIFHGIGSPCYPAOYVGDREWRKYLHLSPALVGLATEELLQV 1206
Db 1494 RYPENNFTGRALARLFHGVSQPNYALVMSRS-NFWRAYTITDFNRIVRLANAEIVRM 1550

RESULT 9
Q9NH11
ID Q9NH11 PRELIMINARY; PRT; 1530 AA.
AC Q9NH11
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE RECQ4.
OS Name=RecQ4; Synonyms=RECQ4;
OC Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Kusano K., Barres M.E., Engels M.R.;
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF233659; AAF2939.1; .
DR HSSP; P15043; 10Y.
DR FlyBase; FBgn0040290; RecQ4.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0016787; F:Hydrolase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006259; P:nucleic acid metabolism; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004589; RecQ.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00098; Zf_CCHC_1.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC_C; 1.
DR SMART; SM00343; ZNF_C2HC; 1.
DR TIGRFAMs; TIGR00614; recQ_fam; 1.
DR PROSITE; PS50158; ZF_CCHC; 1.
KW ATP-binding; Helicase; Hydrolase.
SQ SEQUENCE 1530 AA; 169876 MW; 1B753A59A3B30EB4 CRC64;

Query Match 26.2%; Score 1680.5; DB 2; Length 1530;
Best Local Similarity 32.4%; Pred. No. 8-8e-72;
Matches 420; Conservative 191; Mismatches 412; Indels 275; Gaps 36;

QY 107 ORLKANKLQGLQAGP-ALGRRPWPLGRASSKASTPKPCTGVPSPFAKVSDEPQLBPP 165
Db 306 QELQTDNMSNMWQKQDHLNHTP-PASQSVAPKPKP-----PEQGTDSDDSDVAES 359
QY 166 QPRRLQHLQASLSQRLGSLDPCWLO-QRHSSEVDFLGAKKAC-----RPLDGSSESQ-- 218
Db 360 EEEQEPQRYOLSKRKRTVSTASGVAAAPVEIIPNKVETETFTFAQENPFSADEDDA 419
QY 219 LLIF-----GESAVLDGPAGSQ---GPEASAFQEVISIRVGSQPSGSGEKKRMNEBPWS 271

```

```
Db 420 TVPENKKKAKRKOAGKOKTKPKA-----EPKPKT-----EK 455
QY 272 PAVOOESSQAG--PPSEGAQAVAEDEPPGPVQAOAPPQ-----CSSPSNR 318
Db 456 KAKVAKAEKKPKAEKPKRNSKKAIAVEPADPDEDEKQPLNPDCKVLALEAGDITSVR 515
QY 319 YH-----GLSPSSQARAGKAEGTAPLHIPPRLARHD-RGNY 353
Db 516 INVQDLBEADATAGRYIRTFPAAGPNPGJSEGSNIRIDEBRAAKLEERIAAGKLNENF 575
QY 354 VELANKOKHYVNG-RALRSRLLRKQAWKOKMR-----KKGCGGCGGATVTTKESC 403
Db 576 VVINIOKKKFFVGRKSVNSFKYKQKQOMRHKKRVAAALSGPDMMDGGCDG---VLT---C 628
QY 404 FLNEQPDHWAOC-----PRP----- 419
Db 629 FCCGGVGFPAQCKYKKGDSLPLSLAQLSEDPSPBPPLLAQEMASQAVVAHSRNISRLP 688
QY 420 -----ASEBDTAVGPBPPLVPSBPQ----- 439
Db 689 QANANAILQGDLEINSEEDDQESSSEDEYQCHDPNWSDDMDVDPEALDAAVEASLSQP 748
QY 440 -----VPE---VPSLDPT-----VLPLYSIGPSGQLAEPRA 467
Db 749 VSQEKASPIKTYVGHKIPSEEPKQAGLDTTASSNSRQHGKPKLYDLLPDGSYQDTPR 808
QY 468 EYFOALEQGHQAFRRGGERAVWRLISGSLTLLVPTGAGSLCYQLPPLALYSRRSPCLT 527
Db 809 EYLEALHMFHGTNFRKGQDRALMRTLSGSLVLTSTGSGKSLCYQLPAYLSRKQALIT 868
QY 528 LVVSPSLIMDQVSGLPCLKAACISGNTKQKRESVLOKTRAOVHVLMLPEALV-- 585
Db 869 LVISPLVSLMEQVGVGPHFLRAHCHTNOTAPQRMKIQOMLANGSIDLILVSPBAVNG 928
QY 586 ----GAGGLPPAPOLPPVAPACTIDEAHCLSSWSHNPFCYLRYCKTLRBMGYHCLGIT 641
Db 929 ERATGFGAI--LRQPLPIAFACIDEAHCVSOWSHNRPESYLMICKLRYNLGVRTYLGLT 986
QY 642 ATATRTASDVAGHLAVAE-EPDLHGPARVPNTNLSVMDMDTDQALLTLLQGRPOL 700
Db 987 ATATLPTRVSLTNHGLISDGEKGLISDPLPDNLVLSKDENRDALLQLNSEFEPC 1046
QY 701 DSIIYYCNREDTERIALLRCTCLHAAMPVSGSGRAPKT----TAAVYAHGMSRRRRV 756
Db 1047 QSIIITCTRDCEERLAFIRTCVQDRREPTDQTKKRKRVWQAEPRYAHGAPASRRV 1106
QY 757 QRAFMQGLRVVAVAFGMDRDPDRAVLHLGPRSESVYQAVGARQDQPAHCHL 816
Db 1107 QRAFMNSNEIRIVATIAFGMGINKPDIRAVIHYNMERNPESYVQOIGRAGRDGLPSCHL 1166
QY 817 FLPOGEDRELRLRHVHADSTDFLAVKRLVQRFVPACTCTCTRPSEOGAVGGERPVK 876
Db 1167 FLDAKGGDSELRHHYNSIDRHVTRKLLQKIPVCSG-----DKASAKRTALPIP- 1218
QY 877 YPPOAEQLSHQAPRPR-RVCMGHERALPIQLTVQALDMPBEALITLLCYLELPHNM 935
Db 1219 -----LEBGSFRVHMGSGHIGSVKTYEMDIPRAENITLLCYELHEDPRKCI 1267
QY 936 ELIATTYTCRLNCPGSPRALQALAHRCPLPCLAAQOLPEDP-QGSSSVSEFDMVKVD 994
Db 1268 SVLSAYWMAKVSYSGPKYLKHAACECPPLAMAILQIRDKTFKEDSNLIEFSVDIA 1327
QY 995 SNGWELASVRLALCOLOMDHERPTGVRGTVLVESELSAFLHRSQDLTAEKQOICP 1054
Db 1328 GIGWNSGVVXYQDKDEM--VKVNGYPKSPITVSYDYGFRLYKVGDPTESEIDNALDT 1385
QY 1055 LVGRVQAREFOALARLRTTQAFHSAVAFSCGFCLEQO--DEERSTLXOLLGYYFEEBG 1113
Db 1386 LITRSVKQERTQILQIQYVAHGLAAVAYS CGCCCAADPRDQBGQKAIVRNYFANDYP 1445
QY 1114 Q-----EFGMEDAGPBPGRQARLQDMDQVCDIRQFLSLRPEKFSRAVARLFGHI 1167
Db 1446 QULELEIEPSNVPD-----ENIIDVHALINMYRDNLTFTGRNIRARLFGHI 1490
```

```
QY 1168 GSPCYPAQVGDRRFRFRKYLHLSFHALVGLATEBLLQ 1205
Db 1491 MSPNYPVIMWGR-CKFWRAHYKVDNPENRIHLANMAIILK 1527

RESULT 10
QYR85 ID QYR85 PRELIMINARY; PRT; 340 AA.
AC QYR85;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE RecQ protein-like 4 (Fragment).
GN Name=RECQ4;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14667821; DOI=10.1016/S0888-7543(03)00238-6;
RA Winter A., Alinger A., Fries R.;
RT "Assessment of the gene content of the chromosomal regions flanking
RT bovine DGAT1."
RL Genomics 83:172-180(2004).
RT EMBL: AJ518973; CDS58810.1; -.
DR HSSP; P15043; 10YF.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001650; DEAD.
DR InterPro; IPR002711; Helicase_C.
DR Pfam; PF002711; Helicase_C; 1.
DR SMART; SM00490; HELICC; 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 340 AA; 37366 MW; 205F057DD109CAED CRC64;

Query Match 18.2%; Score 1172; DB 2; Length 340;
Best Local Similarity 67.8%; Pred. No. 2.7e-48;
Matches 232; Conservative 27; Mismatches 57; Indels 26; Gaps 2;

QY 601 FACIDEAHCLSSWSHNPFCYLRYCKV-----LREBMGYHC 636
Db 1 FACIDEAHCLSSWSHNPFCYLRYCVQVSLGKMSVQGRBEGTAGSPPSQTLRDQMGVHC 60
QY 637 FLGLTATATRTASDVAGHLAVAEPPDLHGPARVPTNHLVSMDRDTDOALLTLLQGR 696
Db 61 FLGLTATATRTASDALDMAHNLGVTSESVLRGPATITIDNHLSSSDRDPQALLVTLRSR 120
QY 697 FQNLDSIIYYCNREDTERIALLRCTCLHAAMPVSGSGRAPKTTVAYAHGMSRRRRV 756
Db 121 FALASVLIYYCHRRREDTERVALLRCTCLDADPPHGALEAVEAAYAGLCSRRRRV 180
QY 757 QRAFMQGLRVVAVAFGMDRDPDRAVLHLGPRSESVYQAVGARQDQPAHCHL 816
Db 181 QRAFMEGRRIRRVAVAFGMDRDPDRAVLHLGPRSESVYQAVGARQDQPAHCHL 240
QY 817 FLPOGEDRELRLRHVHADSTDFLAVKRLVQRFVPACTCTCTRPSEOGAVGGERPVK 876
Db 241 FLRPGQDLREKRRRHVHANAVDFAVKRLVQRFV--PCACRQRPEDGSGSESGHLA 298
QY 877 YPPOAEQLSHQAPRPRRVCMGHERALPIQLTVQALDMPBE 918
Db 299 AVVAASADQSGQSPVHTPRCGHERVLPVQPTVQALDMPBE 340

RESULT 11
QYR829 ID QYR829 PRELIMINARY; PRT; 927 AA.
AC QYR829;
```



```

DR HSP: P15043; 10YY.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR GO: GO:0006259; P:DNA metabolism; IEA.
DR Pfam: PF00271; DEAD; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HelicC; 1.
DR TIGRFAMs: TIGR00614; recQ_fam; 1.
DR PROSITE: PS00690; DEAD ATP HELICASE; UNKNOWN_1.
DR ATP-binding; Helicase; Hydrolase.
SEQUENCE 870 AA; 96382 MW; 8BSADA4F7A81723 CRC64;

Query Match      14.0%; Score 897.5; DB 2; Length 870;
Best Local Similarity 26.7%; Pred. No. 8.6e-35;
Matches 295; Conservative 160; Mismatches 368; Indels 281; Gaps 38;

QY 133 ASSKATPKPGGVPVPAEKVSDERPDLPERPGRLOHQLQASLSORLSLDPMQ 192
DB 3 SSSSSSKPRAP--THPPNPQAEAVPSPYPPPPPS-----LFTNLPRIQSDPA--- 53
QY 193 RCHSEVPDLGAPKACRPDLGSEESQLIPGSAVLPGAGSGQPEASAFQEVSRVQSP 252
DB 54 RFSSSVSF---SRLCRSASFTSVEKLKSDGVDFV-----PEPLVEVIA----- 95
QY 253 QSSSGGKKRRNNEEPWESPAVOQESSQAGPPSBGAVAVEBPPG---EPVQAQPPQ 309
DB 96 -----PPK-----SVRRKPPNITPTITSPPK 118
QY 310 PCSSPENRPHVGLSPSSQABAGAEPTAHLTFPRLAHRDGNVYLNKKQHYVGRAL 369
DB 119 PWFPSN-----GNGE-----GNFVGLNNGK---RGKRF 145
QY 370 RSRL-----LRQAWKQKMKKGCFCGGATVTTKSCFNEQFDHWAACPPRASEE 423
DB 146 PKKYGVSKRSRYSRGRKRYKKKENDGGESELL--EEESDQKQ-----ED 191
QY 424 DTDANGPEPLVSPQCPVEVPSIDPTVPLYSIGSGQLAEPPAEVQALBOLGHOAPR 483
DB 192 EANGF-----ISSVEDAIIAIVKTEASDENITKILNIVY-----GYDSFRD 231
QY 484 GGBAIVMRLSGISTLLVPTGAGSKLCYQALPALTYSRSPCLTIIVPSLLSIMDDVSG 543
DB 232 GQLQATKMLGSSSTLVPTGAGSKLCYQIPMTIL---PGTIVASPLVSLMIDQLN 287
QY 544 LPPLCLACIHSQMTKQRESVLQKIRAAQVHMLTPALVAGAGLPAAQLPVAFAFC 603
DB 288 LPSITIGGLSSQREBEATELTRKKEGIKVLFPVSPRLINVEFLSMFRSLSVLV 347
QY 604 IDBAHCLGWSHNPFCYLRV-CKVIREMGVHCFLGTATATRTPTASVAOHLVAEER 662
DB 348 VDEAHCVSESHNFRSPYMKLAKASMLFSELKACILAMTATATMTLQAVMSLEI-PST 406
QY 663 DLHGPAVPVNTNLISVM---DRDTPQALLTLTGKRPQNLISIIYCCRRRETEIAL 719
DB 407 NLIQSQLDNDELISLISGANMKD--LILIMESPVEIKSIIYCYQFETDITISKY 464
QY 720 LFTCLAAAVPGSGRAPKTTAEAYAHAGCSRERRVQRAFQMGQLRVVAVATVAFMGMD 779
DB 465 LND-----NNINAKGYSHGLPAKDRVAIQSFSCNKRIVVAVATVAFMGMD 510
QY 780 RPDVRAVHLGLPPESESYVQAVGARGDQPAHCHLFLQPOGEDILRELRRHVADSTP 839
DB 511 KGVGVAVIHFSVYGSEWEYVQEIGRAGDRLSYCHLFY--DNDTVLKRLSHASDVDE 568
QY 840 LAVKRLVGVFRACCTCTRPSEBGAGVGERPVKYPQAEQSLSHQAAQPRRVCMG 899
DB 569 YAVGKFLTVF-----STETKQHR-----KTC-- 590
QY 900 HRAALPIQLVQALDMPBEAIFTLICYLEIHPHMLLATTATYTHGRNLCPGGPQLQAL 959

```

```

DB 591 ---SLVIESAQKEDPMKEEVNQITTLHLEGEVQYLRMLPOLNICTLN----- 636
QY 960 AHRCEPLAVCLAQQLPEBPGGSSVSEFDMKLVDSWG---WELASVRAALCQLQWDEHP 1016
DB 637 FHKSSPNTLAAR-----SALVAAILKSHVQGLHVPDIPAVASSICVAT----- 681
QY 1017 RTGVARGTVLVEFSELAFLHSPDGLTAEEKD-----QICDPLGVRQABER 1064
DB 682 -----IDVLAELQIA-----LKMGEVITYELKDSAFCYTILKSPKEICSLSHLTKMLTE 730
QY 1065 QALARLR-TFOAFHSAVPSCGCPLEQDEBERSTRKOLLGRYFEEBEGEPGGMEDAQ 1123
DB 731 IESCVRKLDIMSSAAVAIAISVNTSELSSGAKQTR--SLQSRIFDYDNG-----DBK 781
QY 1124 GPEPGQARLQWEDQ-----VPCDTRQFLSLRPEEKSSRAVARIPIHGIGSPCAQVYGO 1179
DB 782 CDSPSKA-----TQNCAPLRADIVFPLQSNRQAKFTRALRIKIHGVGSFAPFNVMSK 835
QY 1180 DRFRKRYLHLSFHVLGLATEEL 1203
DB 836 -THFWGRYMNVDFRVIMEAAQTEL 858

RESULT 13
O9C6N0
ID O9C6N0; PRELIMINARY; PRT; 941 AA.
AC O9C6N0;
DT 01-UN-2001 (TRENDEL 17, Created)
DT 01-UN-2001 (TRENDEL 17, Last sequence update)
DT 01-MAR-2004 (TRENDEL 26, Last annotation update)
DE ATP-dependent DNA helicase RecQ, putative.
GN Name=F28L5.4;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
RA Maiti R., Romling C.M., Koo H., Fujii C.Y., Uterback T.R.,
RA Barnstead M.E., Bowman C.L., White O., Mierman W.C., Frazer C.M.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC079280; AAG50580.1; -.
DR HSP: P15043; 10YY.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR GO: GO:0006259; P:DNA metabolism; IEA.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR011545; DEAD/DEAH_N.
DR InterPro: IPR002464; DEAH_box.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR004589; RecQ.
DR Pfam: PF00271; Helicase_C; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HelicC; 1.
DR TIGRFAMs: TIGR00614; recQ_fam; 1.
DR PROSITE: PS00690; DEAD ATP HELICASE; UNKNOWN_1.
DR ATP-binding; Helicase; Hydrolase.
SEQUENCE 941 AA; 104155 MW; 30BC454C795522E1 CRC64;

Query Match      13.6%; Score 872; DB 2; Length 941;
Best Local Similarity 26.2%; Pred. No. 1.5e-33;
Matches 305; Conservative 165; Mismatches 390; Indels 306; Gaps 43;

QY 93 TGRGRGQSVPRYGGRLANKGTIQAGRALGRRPWPLGRASSKSTPKPGTGVPSFA 152
DB 15 TPRD---SPSPSPQLQSPAHV---PYVSRKM---TSSSSSSKPRAP--THPPNVS 62
QY 153 EKVSDPEQLPEPQRPRLQHLQASLSORLSLDPMQORCHSEVPDLGAPKACRPDL 212

```

```

Db 63 QBAVPSPYPPPPPPSP-----LFTNLPRICOSQPA---RFSSVSF---SRLCSRAS 111
Qy 213 GSEBSOLLPGBSAVLPGAGSGCPGASAFQFVSIVGSPQSSGCKRNMEEWESP 272
Db 112 FTSVEKLKSDGVDFV-----PPPLVEVIA----- 136
Qy 273 AVOOESSQAGPPSEGAGAVAEEDPPG---EPVQAPPOPCSSPSPNRYHGLSPSSQAR 329
Db 137 -----PPK-----SVRRKPMLITDITSPYKPMVFRN----- 166
Qy 330 AGKAESTADLHIFRLARHGRNYVLRNKKQHYVGRALRSFL-----LRQAWKQKM 383
Db 167 -NGGE-----GNFVLTNLNGK---RGKFPSPKXKGVSKSRSSSPRGKR 206
Qy 384 RKKGECFGGCAVTTKESCFLNEQPDHMAQCPRASEEDDVAPEPELVSPQVPEV 443
Db 207 YKKKEADGGESL--EESSDLOKOT-----BDEANGF-----I 238
Qy 444 PSLDPTVLPLYSIPSGQLAETPAEYFOALEQHPAPRGQERAVMRLISGISTLVLP 503
Db 239 SSVEDAILAVKTEASDENLTKLNLVY-----GYSPFDGQALIKMLGGSTMLVLP 292
Qy 504 TGAGSLCYQPLALLYSRRSPCLTLVSPPLSLMDQVSGLPCLKAACISGKMKORE 563
Db 293 TGAGKSLCYQIPAMIL---PGITLVSPVLSIMIDQLKHLPSIIGGLSSQPREAT 348
Qy 564 SYLQKIRAAQVAVMLTPRALVAGAGLPPAAQLPPAFACIDBAHLSQ----- 612
Db 349 ETLRLKKEGIIKVLFPVSPERLNVFELSMFRMSLSVLVAVDEAHVCSEICLSLFCIS 408
Qy 613 W-----SHNRPCTYLRV--CVLYBERMGVHCFGLTATARTRTSDVAOHLVAEED 663
Db 409 WILRALIRNSHFRBSYMLKASMLFSELKACILMTATMTITQAVMSLIEI-PSYN 467
Qy 664 LHGPAVPPTNLHLSVSM---DRDTPQALITLQCKRFQMLDSIIYCNREDTERIALL 720
Db 468 LQKQOLRNFELSVLSGANMKD--LILMESPPYKEIRSIYVCKQYETDMISKYL 525
Qy 721 RCLHAAMVPGSGGAPKTTAAEYAHAGMSRERRRQYAFMGQLRVVAVATAFGKGLDR 780
Db 526 RD-----NNINAGYHSGLPKADKRVRIQESFCSNKIRVVAVATAVAFMGGLDK 571
Qy 781 PVRAVILHLGLPPSFESYQAVAGRAGDOPAHCHFLDPQGEDRLRELRHNAADTFL 840
Db 572 GIVGAVIHSVPGSHEEIVQEIGRAGDRDLSTCHLFLY--DNDTYKILSLASHSDVDY 629
Qy 841 AVKRLVQVFPACTCTCTRPSPSEOGAVGGERPVKYPQOEAQOLSHQAPGRVRCMGH 900
Db 630 AVGKFLTHF-----STETKQHE-----KIC-- 650
Qy 901 EBALPIQLTVQALDMEBAIEFLLCYLELHPHHMIELLATYTHCRNLCPGPAQLOALA 960
Db 651 --SLVIESASQKFDMEEWQITLITLHELGEVOYLMLPOLNICCTLN-----F 697
Qy 961 HR---CPPLAVCLAQOLPEDPGQSSSVFDMVKLVDSWG---WELASFRALCOLQMDH 1014
Db 698 HTTSDAQNLVMLYVSSPNTLAARSATVAIILKSHVKGGLHVFDPAAVASSICVAT-- 754
Qy 1015 EPRTRGRTGVLVSESLAFHLRSPGDLTAEBKD-----QICDLYLGVQAR 1062
Db 755 -----TDVLAELQNL-----KGEVTEILKQSAFCYITLKSKEICLSLSHLTKML 799
Qy 1063 EBOALARLR--TFQAFHSVAFPSGCPCLQOQDEBSTRKLDLGRYFEEBEOQEPGMD 1121
Db 800 TEIESKVKAKLDIMSSAAVAIVSNTSELSGAKQTR--SLQSRIFDYFG-----D 850
Qy 1122 AGSPBGQARLDWEDQ-----TRCDIROFLSLRPEKFSRAVARIFHIGSPCYAOVY 1177
Db 851 EKCDSBSKA-----TONCAFRLADIKVFLQSNRQAKFTPRAIARIMHGSSPAPFNSVW 904
Qy 1178 GQDRRFWRKYHLHSFHALVGLATEEL 1203
Db 905 SK-THFMGRYMWVDFRVIMEAQOTEL 929

```

```

RESULT 14
Q84RP1 ID Q84RP1 PRELIMINARY; PRT; 899 AA.
AC Q84RP1;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Similar to ATP-dependent DNA helicase.
GN Name=H0302E05.4;
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhacoidae; Oryzae; Oryza.
OX NCBI_TaxId=4530;
RN (1)
RP SEQUENCE FROM N.A.
RX PubMed=14693376; DOI=10.1016/j.gene.2003.09.008;
RA Wang R., Hong G.F., Han B.;
RT "Transcript abundance of rml1, a putative GFI-like factor in rice, is
RT up-regulated by Magnaporthe grisea and down-regulated by light.";
RL Gene 324:105-115(2004).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=22780373; PubMed=12897975;
RA Wang R., Hong G.F., Han B.;
RT "Characterization of the copy number of RIR10 retrotransposon and
RT analyses of transcriptional activity of its LTR in rice genome.";
RL Acta Biochim. Biophys. Sin. 35:768-773(2003).
DR EMBL; AL627350; CAD79701.1; -.
DR HSSP; P15043; 10YV.
DR Gramene; Q84RP1; -.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0006259; P:DNA metabolism; IEA.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR004589; RecQ.
DR Pfam; PF00270; DEAD_1.
DR SMART; SM00487; DEXDC; 1.
DR TIGRFAMs; TIGR00614; recQ_fam; 1.
KW Helicase.
SQ SEQUENCE 899 AA; 99340 MW; 48D6A43F055749C CRC64;

Query Match 12.0%; Score 772; DB 2; Length 899;
Best Local Similarity 23.8%; Pred. No. 8.3e-29;
Matches 286; Conservative 165; Mismatches 390; Indels 362; Gaps 41;

Qy 40 ALYREYRLKRTTGGAGGGLSSSESLPAAABEAPRCMGPHLNAATKSPQTPGRGRQ 99
Db 16 AFFLARLRIRCTPPAHAAADSDSGSLSDVSASPPRRSP--LRPAPPPPPPPKGR- 72
Qy 100 GSVVDYGRKLKMLKGTLLQAGPALGRRPWPLGRASAKSTPKRPCTGVPVSPFAKVSDEP 159
Db 73 -----TRDAAPTKPLKKTTPAASAPAPAP 97
Qy 160 POLPEPPRPRGLQLQASLSQRLGSLDP--GMLQCHSEVPDFLQAPKACRPDLCSSES- 217
Db 98 P--PPPTLR-----AAALS-----DPHGIAAR-----IAAGSALTAASGTASS 134
Qy 218 --QLIIPESAVLPGAGSGCPGASAFQFVSIVGSPQSSGCKRNMEEWESPAPQV 275
Db 135 SFRRLVQSRNPSFDPATFTAPASABSEVSAARPPPTAA-----TAPROT 183
Qy 276 QOESSQAGPPSEGAGAVAEEDPPGEPVQAOPOPCCSSPNRNYHGLSPSSQARAKAEG 335
Db 184 RPKRVHPSVSEVAAASAABQP-----KXARG----- 212
Qy 336 TAPLHIFRLARHGRGNYVLRNKKQHYVGRALRSRLIRQAWKQKRGKGCSCGGA 395
Db 213 -----SEGNFVRUNI--NGYGRRTFKNSQAKSTKCRSWRQ---RAAGA 253
Qy 396 TVTTSKCFLENGPDHMAQCPRASEEDDVAPEPELVSPQVPEVPSLDPTVLPLYS 455

```

```

Db 254 T-----PRSQDEBEGDLVAEALLEREKOASD-----SVLEAVE 287
QY 456 LGPSQGLAETPAVEFOALEQ-----GHOAFPPGGERAVMRILSGISTLLVPTGAGKSL 510
Db 288 -----SVEDEPE--QNLKSLINAAVGHDSFROGQLEAIQOIVAGBSTLVLPFGAGNSL 340
QY 511 CYOLPALLYSRSPCLITLVVSPPLSLMDQVSGLPCCAKACIHSGMTKORESVLOKIR 570
Db 341 CYQVPAAML-----PGLTLVSPPLSLMDQVQLKRLPAFLPGGLASSOTSDEBHTLQRLR 396
QY 571 AAQVAVMLTPRALVAGAGLPRAAQLPPVAFACIDEAHLCSWSHNFRCYLR--CKVLR 629
Db 397 AGEIVLFPVSPERFLNEBELLFRDTLPISLVAIDEAHICISMSHNFRCYLRASLIR 456
QY 630 EBMGVHCFGLTATRTASDVAAQHLAAVEEPDLHGPAVPNTNLISMDRDTDOALL 689
Db 457 RKLAVQCLAMTATVTTOTLEIEMNALEIPSD-NLIQTSQIENIQLSISTDNRLKDIM 515
QY 630 TLLQGRFPQNDLSIIYCNRRREDTERIALRLTCLHAAMVPGSGRAPRTTAAVHAAGMC 749
Db 516 LLLKSPFPVDMRSIIYCKFQAEITDPSKYL--CDN-----NITAKSYHSGLL 561
QY 750 SERRRVQRAFMQGLRVVAVTAFAKGLDRPDVAVLHLGLPPSFESYQAVAGARDG 809
Db 562 IKRRSRVDELFCSSNKRIRP-----SY----- 582
QY 810 QPAHCHLFLQPGEDRLERHVAHSDTDFLAVKLVORVPFACCTCTPRPSBOGAVG 869
Db 583 -----NFTBOP-----FLGSDGVGYAMSKFLYQIF-----SBENTTG 615
QY 870 GERPVKYPPOBAEQLSHQAPRGPRVCMGHERALPIQLTVALDMPBEAIEITLYCYEL 929
Db 616 -----FTHKTSFQILADKDLIRSVLNSSEMDGHY 694
QY 930 HPHHMLLATTYTHCRCLNCPGPAQOLALARCPLAVC-----LAQQLPEPPGQSSSV 985
Db 646 GDOQVIRLLPQPSVCTL-----YFHKTSFQILADKDLIRSVLNSSEMDGHY 694
QY 966 EFDPMVLVDSMGWELASVBRALCQLOMDHEPRTVARGVGLVBPSS-ELAFHLRSPG--- 1041
Db 695 VEDIRIANDLKITMNEV-----FDHLHK-----LKSGLSIFLKDPAICY 736
QY 1042 -----DLTAEKQICDPLVGRVQARERQALRLRTFOAFHSVAFPSG----- 1085
Db 737 VILMRPDDNALSAN-----LTKMLSEVSSKISKLDAWF--ALANFAVGCRRTGSCSS 790
QY 1086 ---GPCLEQODBERSRLKDLGRYFEESGQEPGCMEDAQGPBGQARLQDMEDQVRCI 1143
Db 791 QHTPCIOKKIME-----YFSKDDGTSEBNC-----RTQLQKSPFLQADI 830
QY 1144 RQFLSLRPEKFSRAVARIIPHIGSPCYPAQVGGDRPFMRKYHLISFHALVGLATEL 1203
Db 831 KVFQISNSAKFTPRVARIIMHGISPAFPVTSWKN-HFMGRYVEVDPLVWEAKKAE 889
QY 1204 LQV 1206
Db 890 VKL 892

```

```

RESULT 15
07FAH0 PRELIMINARY, PRT, 874 AA.
AC 07FAH0;
DT 05-JUL-2004 (TREMblrel. 27, Created)
DT 05-JUL-2004 (TREMblrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMblrel. 27, Last annotation update)
DE OJ000223 09.17 protein.
GN Name=OJ000223 09.17;
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxId=39947;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12447439; DOI=10.1038/nature01183;
RA Peng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Jing K., Yu S., Tang Y.,
RA Wang Q., Zhang L., Lu Y., Mu J., Lu Y., Zhang L.S., Yu Z., Fan D.,
RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,
RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
RA Han B.,
RT "Sequence and analysis of rice chromosome 4.";
RL Nature 420:316-320(2002).
DR EMBL; AL606998; CAD41405.2; -.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0006259; F:DNA metabolism; IEA.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR004589; RecQ.
DR Pfam; PF00270; DEAD; 1.
DR SMART; SM00487; DEXDC; 1.
DR TIGRfam; TIGR00614; recQ_fam; 1.
SQ SEQUENCE 874 AA; 96274 MW; 51B64C486617774F CRC64;

Query Match 11.8%; Score 761; DB 2; Length 874;
Best Local Similarity 23.8%; Pred. No. 2,7e-28;
Matches 271; Conservative 163; Mismatches 377; Indels 330; Gaps 38;

110 KANLKTGLQAGPALGRPWPGLGRASSKASTKPP--GTGVPVSPAEKVSDBEPOLPEBQ 166
Db 13 EGSILSDVASAPRRRSP--PRPAPPPPPPPPKHTRPAAPKPKIKPPTPAASAPA 69
QY 167 PRPGRLQLQASLSRGLSLDP-GWLQCRHSVVPFLDAPACRDLDSESS---QLILP 222
Db 70 PAPPPTLRLAAS-----DPHGLAR-----IAGSLTLTASGTAASSSRRLVQ 116
QY 223 GESAVLPGAGSGCGEASAFQESVIRVGSPPSSSGEKRRNNEBPWSPAQVQSSQA 282
Db 117 SRNPSFDATNFTASASSAPSEVPSAARPPETA-----TDAPQTRKRVAP 165
QY 283 GPSSGAGAVAVEPDPGPVQAOQPPQCSPPSNRYHGLSPSSQARAGKAGTAPLHF 342
Db 166 NSVSEVAAASAAAEQP-----KVARGG----- 187
QY 343 PRLAHDGRGVYRLMMQKHVYGRALRSRLRKQAMKQKKKCEGCGGATYTKES 402
Db 188 -----SGNFFRLNI--NGTGRRTTFKNSQAKSTKCSMRKQ---RAAGAT----- 229
QY 403 CFILNQPDHMAAQCPRPASBEDTDAVGPEPLVPSQVPEVPSLDPTVLPLYSLGPSQL 462
Db 230 -----PRSGDEBEGDLVAEALLEREKOASD-----SVLEAVE-----SV 264
QY 463 AETPAVEFOALEQ-----GHOAFPPGGERAVMRILSGISTLLVPTGAGSLCYQAL 517
Db 265 REDPSE--QNLKSLINAAVGHDSFROGQLEAIQOIVAGBSTLVLPFGAGNSLQVPA 322
QY 518 LYSRSPCLITLVVSPPLSLMDQVSGLPCCAKACIHSGMTKORESVLOKIRAAQVAVL 577
Db 323 IL-----PGLTLVSPPLSLMDQVQLKRLPAFLPGGLASSOTSDEBHTLQRLRAEIKVL 378
QY 578 MLTPRALVAGAGLPRAAQLPPVAFACIDEAHLCSWSHNFRCYLRV--CKVLRBMGVHC 636
Db 379 FVSPERFLNEBELLFRDTLPISLVAIDEAHICISMSHNFRCYLRASLIRKLAVQC 438
QY 637 FLGLTATVTRRTASDVAAQHLAAVEPDLHGPAVPNTNLISVMDRDTDOALLTLQGR 696
Db 439 ILAMTATVTTOTLEIEMNALEIPSD-NLIQTSQIENIQLSISTDNRLKDIMLLKSP 497
QY 697 PQNDLSIIYCNRRREDTERIALRLTCLHAAMVPGSGRAPRTTAAVHAAGCSERRRV 756
Db 498 FVDMRSIIYCKFQAEITDPSKYL--CDN-----NITAKSYHSGLLIKRSKV 543

```

```
QY 757 QRAFMGOLRVVATVAFMGDLRDPRAVLHLGLPSPESYVQAVGRAGRDQPAHCHL 816
| : : |
| : : |
Db 544 QELFCSNKIRVF-----SY-----NF 559
| : : |
QY 817 FLQPGEDLRELRRHVHADSTDFLAVKRLVQRFPA-----CTCTRPPSBOGAVGE 871
| : : |
| : : |
Db 560 TEOP-----FIGSDGVGYAMSKFLYOIFSESENTTGCISLAK-----597
| : : |
QY 872 RVPVKYPPQEAQLSHQAAFGPRRCVMGHERALPIQLTVQALDMPBEALETLLCYLELHP 931
| : : |
| : : |
Db 598 -----ELTSRKPDIKEEVLLTTLTOLLEIGD 622
| : : |
QY 932 HHMLRLATTTTHCRINCPGPAQLOALAHRCPLAVC---LAQOLPEDPGSSSVVF 987
| : : |
| : : |
Db 623 QOYIRLLPQFSYCTL-----YHKTSPLQIADKDILIRSVLNRSBMDGHYVF 671
| : : |
QY 988 DMVKLVDSKMGWELASVRALCOLMDHEBRTGVRGTGVLVEFS-ELAFHLRSPG-----1041
| : : |
| : : |
Db 672 DIPRIANDLKITMNEV-----FDHLHK-----LKFSGEISFELKDPAVCYVI 713
| : : |
QY 1042 -----DLTAEKDQICDPLYGKVGQARERQALRLRTFOAFHSVAFPSG-----1085
| : : |
| : : |
Db 714 LWRPDDFNALSN-----LTKWLSVESSKISKLDAMF-ALANFAVKGCKRTGCGSGSQH 767
| : : |
QY 1086 GPCLBQDEBERSTRKDLGRYFEBEEOGEPGMEDAQGPBQARLODMEDQVRCDIRQ 1145
| : : |
| : : |
Db 768 TPCIOKKIME-----YFSKDDGTSENDG-----RTQLOKSSPFLQNDIKV 807
| : : |
QY 1146 FLSLRPEEKFSRAVARIFPHIGISPCYPQAVYGODRRFRKYLHLSPHALVGIATEELLQ 1205
| : : |
| : : |
Db 808 FIOSNSPAKFTPRAVARIMHGISSPAFAFVTSKN-HFWGRYVEVDPLVMEAKAKELVK 866
| : : |
QY 1206 V 1206
| : : |
Db 867 L 867
```

Search completed: March 19, 2005, 01:08:06
Job time : 210 secs

QY 301 EBYOAPQPCSSPSNPRYHGLSPSSQABGAEGTAPLHIIPRLARHDNGYVNLNMKQ 360
 DB 301 EBYOAPQPCSSPSNPRYHGLSPSSQABGAEGTAPLHIIPRLARHDNGYVNLNMKQ 360
 QY 361 KHYVGRALRSRLRKQAMKQKRRKKGECFGGGGAVTTTKESCFLNEQFDHMAAOCPRPA 420
 DB 361 KHYVGRALRSRLRKQAMKQKRRKKGECFGGGGAVTTTKESCFLNEQFDHMAAOCPRPA 420
 QY 421 SEEDTAVGPEPLVPSPOVPEVPSLDPTVLPLYSIGPSGQLAETPAEFOALEQGHQA 480
 DB 421 SEEDTAVGPEPLVPSPOVPEVPSLDPTVLPLYSIGPSGQLAETPAEFOALEQGHQA 480
 QY 481 FRPGERAVMRILISGISTLLVLP.TGAGKSLCYQLPALLYSRSPCLTLVSPLSLMDQ 540
 DB 481 FRPGERAVMRILISGISTLLVLP.TGAGKSLCYQLPALLYSRSPCLTLVSPLSLMDQ 540
 QY 541 VSGLPCLKAACHSGMTKQRESVLOKIRAAQVHVLMTPEALVAGGLPPAAQLPVA 600
 DB 541 VSGLPCLKAACHSGMTKQRESVLOKIRAAQVHVLMTPEALVAGGLPPAAQLPVA 600
 QY 601 FACIDBAHCLSQWSNFRPCYLKVBKVBGMVHCFGLTATATRTASDVAOHLAVAE 660
 DB 601 FACIDBAHCLSQWSNFRPCYLKVBKVBGMVHCFGLTATATRTASDVAOHLAVAE 660
 QY 661 EPDLGPAVPPTNLHLSVSMRDTQALLTLLOGRFQNLDSIIICNRREDTERIAALL 720
 DB 661 EPDLGPAVPPTNLHLSVSMRDTQALLTLLOGRFQNLDSIIICNRREDTERIAALL 720
 QY 721 RCTLHAAMVPGSGGAPKTTAEYHAGMCSRRERRVQAPMGOQLVAVVATVAFGKGLR 780
 DB 721 RCTLHAAMVPGSGGAPKTTAEYHAGMCSRRERRVQAPMGOQLVAVVATVAFGKGLR 780
 QY 781 PDVRAVLHGLPSPFSFYQVAVRAGRDQPAHCHFLQPOGEDLRELRHYADSTDFL 840
 DB 781 PDVRAVLHGLPSPFSFYQVAVRAGRDQPAHCHFLQPOGEDLRELRHYADSTDFL 840
 QY 841 AVKRLVQVFPACTCTCTRPSPBQEGAVGGERVPKYPQEBABQLSHQAAPGRVCMGH 900
 DB 841 AVKRLVQVFPACTCTCTRPSPBQEGAVGGERVPKYPQEBABQLSHQAAPGRVCMGH 900
 QY 901 ERLALPQLTVQALDMEBEALETLLCYLELHPHMLLATTYTHCGLNCPGSAQOQALA 960
 DB 901 ERLALPQLTVQALDMEBEALETLLCYLELHPHMLLATTYTHCGLNCPGSAQOQALA 960
 QY 961 HRCPLAVCLAQOLPEDPCGSSSVFEDWVKLVDSMGWELASVRALCOLQMDHEPRTGY 1020
 DB 961 HRCPLAVCLAQOLPEDPCGSSSVFEDWVKLVDSMGWELASVRALCOLQMDHEPRTGY 1020
 QY 1021 RRGTVLVEFSSELAFLHRSBGDLTAEBKQOICDFLYGRVQABERQALARLRTEQA.FHSV 1080
 DB 1021 RRGTVLVEFSSELAFLHRSBGDLTAEBKQOICDFLYGRVQABERQALARLRTEQA.FHSV 1080
 QY 1081 APPSCPCLEQODEESTRLKOLLGRYFEEBEOEGEGMEDAOPGPGQARLODMWDQVR 1140
 DB 1081 APPSCPCLEQODEESTRLKOLLGRYFEEBEOEGEGMEDAOPGPGQARLODMWDQVR 1140
 QY 1141 CDRQCLSLRPEKFFSSRAVARIFHGIGSPCYPAQYVYODRRFMRKYLHLSPALVGLAT 1200
 DB 1141 CDRQCLSLRPEKFFSSRAVARIFHGIGSPCYPAQYVYODRRFMRKYLHLSPALVGLAT 1200
 QY 1201 EBLLOVAR 1208
 DB 1201 EBLLOVAR 1208

RESULT 2
US-09-699-135-2

; Sequence 2, Application US/09699135

; Patent No. 6472513

; GENERAL INFORMATION:

; APPLICANT: AGENE Research Institute, Co., Ltd.

; APPLICANT: HIRAKI AND ASSOCIATES

; APPLICANT: SHIMAMOTO, AKIRO

; APPLICANT: KITAO, SAORI
 ; APPLICANT: FURUICHI, YASUHIRO
 ; TITLE OF INVENTION: HUMAN GENE SEQ4 ENCODING HELICASE
 ; FILE REFERENCE: HIRAL150
 ; CURRENT APPLICATION NUMBER: US/09/699,135
 ; PRIOR FILING DATE: 2000-10-27
 ; PRIOR APPLICATION NUMBER: US/09/463,702A
 ; PRIOR FILING DATE: 2000-01-24
 ; PRIOR APPLICATION NUMBER: PCT/JP98/03114
 ; PRIOR FILING DATE: 1998-07-10
 ; PRIOR APPLICATION NUMBER: JAPAN 9/200387
 ; PRIOR FILING DATE: 1997-07-25
 ; NUMBER OF SEQ ID NOS: 44
 ; SOFTWARE: Patencin version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 1208
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-699-135-2

Query Match 100.0%; Score 6424; DB 4; Length 1208;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MERLDVRERLQAMERAPRRQRRPSQDDVAAPEETRALYREYRTLKRTTGAAGGIR 60
 DB 1 MERLDVRERLQAMERAPRRQRRPSQDDVAAPEETRALYREYRTLKRTTGAAGGIR 60
 QY 61 SEESLPAAAEAPERCWCPHLNRAATKSPOTPPERSQGSVPDQGRKANLKGTLQNG 120
 DB 61 SEESLPAAAEAPERCWCPHLNRAATKSPOTPPERSQGSVPDQGRKANLKGTLQNG 120
 QY 121 PALGRRPMPFLGASASKASTPKPGTGPVPSFAEKYSDEBPQLEPQPPRGRLOHLQASIS 180
 DB 121 PALGRRPMPFLGASASKASTPKPGTGPVPSFAEKYSDEBPQLEPQPPRGRLOHLQASIS 180
 QY 181 QRLGSLDPGWLQRCHEVVDPLGAAPKACPDIGSEBSQLLIFGEZAVLGPAGSGQPEAS 240
 DB 181 QRLGSLDPGWLQRCHEVVDPLGAAPKACPDIGSEBSQLLIFGEZAVLGPAGSGQPEAS 240
 QY 241 AFOEVSIRGSPQSPSSGGEKRMNEEPWESPAOVQOESSQAGPSEGAVAVEDPPG 300
 DB 241 AFOEVSIRGSPQSPSSGGEKRMNEEPWESPAOVQOESSQAGPSEGAVAVEDPPG 300
 QY 301 EBYOAPQPCSSPSNPRYHGLSPSSQABGAEGTAPLHIIPRLARHDNGYVNLNMKQ 360
 DB 301 EBYOAPQPCSSPSNPRYHGLSPSSQABGAEGTAPLHIIPRLARHDNGYVNLNMKQ 360
 QY 361 KHYVGRALRSRLRKQAMKQKRRKKGECFGGGAVTTTKESCFLNEQFDHMAAOCPRPA 420
 DB 361 KHYVGRALRSRLRKQAMKQKRRKKGECFGGGAVTTTKESCFLNEQFDHMAAOCPRPA 420
 QY 421 SEEDTAVGPEPLVPSPOVPEVPSLDPTVLPLYSIGPSGQLAETPAEFOALEQGHQA 480
 DB 421 SEEDTAVGPEPLVPSPOVPEVPSLDPTVLPLYSIGPSGQLAETPAEFOALEQGHQA 480
 QY 481 FRPGERAVMRILISGISTLLVLP.TGAGKSLCYQLPALLYSRSPCLTLVSPLSLMDQ 540
 DB 481 FRPGERAVMRILISGISTLLVLP.TGAGKSLCYQLPALLYSRSPCLTLVSPLSLMDQ 540
 QY 541 VSGLPCLKAACHSGMTKQRESVLOKIRAAQVHVLMTPEALVAGGLPPAAQLPVA 600
 DB 541 VSGLPCLKAACHSGMTKQRESVLOKIRAAQVHVLMTPEALVAGGLPPAAQLPVA 600
 QY 601 FACIDBAHCLSQWSNFRPCYLKVBKVBGMVHCFGLTATATRTASDVAOHLAVAE 660
 DB 601 FACIDBAHCLSQWSNFRPCYLKVBKVBGMVHCFGLTATATRTASDVAOHLAVAE 660
 QY 661 EPDLGPAVPPTNLHLSVSMRDTQALLTLLOGRFQNLDSIIICNRREDTERIAALL 720
 DB 661 EPDLGPAVPPTNLHLSVSMRDTQALLTLLOGRFQNLDSIIICNRREDTERIAALL 720
 QY 721 RCTLHAAMVPGSGGAPKTTAEYHAGMCSRRERRVQAPMGOQLVAVVATVAFGKGLR 780

DB 721 RTCHAAVPPSGGAPKTTAAVHAAGCSRRRRVQAFMGOGLRVVATAVAFMGGLDR 780
QY 781 PDVRAVHLGLPSPFESYVQAVGRAGRGOPAHCHFLPOPGEDLRELRHHVADSTDFL 840
DB 781 PDVRAVHLGLPSPFESYVQAVGRAGRGOPAHCHFLPOPGEDLRELRHHVADSTDFL 840
QY 841 AVKRLVORVFPACTCTCRPPSEOGAVGGERPVPKYPOEAEOLSHOAPGPRVCMGH 900
DB 841 AVKRLVORVFPACTCTCRPPSEOGAVGGERPVPKYPOEAEOLSHOAPGPRVCMGH 900
QY 901 BRALPIQTLVQALDMPBEAIEITLICYELHPHMLLATTYTHCLNCPGPAQLOALA 960
DB 901 BRALPIQTLVQALDMPBEAIEITLICYELHPHMLLATTYTHCLNCPGPAQLOALA 960
QY 961 HRCPLAVCLAQOLPBDPGGSSSVFEDMKLVDSMGWELASVRALCOLQMDHEBRTGV 1020
DB 961 HRCPLAVCLAQOLPBDPGGSSSVFEDMKLVDSMGWELASVRALCOLQMDHEBRTGV 1020
QY 1021 RRGTVLVFSESLAFHLSRGDLTAEBKQICDPLVGRVQAREGRLARTRTFOAFHSV 1080
DB 1021 RRGTVLVFSESLAFHLSRGDLTAEBKQICDPLVGRVQAREGRLARTRTFOAFHSV 1080
QY 1081 AFPSCGPLEQODEBERSTRKDLGRYFEEBEGOEFGMEDAQGPQARLQDMEDQVR 1140
DB 1081 AFPSCGPLEQODEBERSTRKDLGRYFEEBEGOEFGMEDAQGPQARLQDMEDQVR 1140
QY 1141 CDIROPFLRPEKFSRAVARIFHIGISPCYPAQVYGDRFRFKYHLSPHALVGLAT 1200
DB 1141 CDIROPFLRPEKFSRAVARIFHIGISPCYPAQVYGDRFRFKYHLSPHALVGLAT 1200
QY 1201 BELLQVAR 1208
DB 1201 BELLQVAR 1208

RESULT 3
US-09-463-702A-37
; Sequence 37, Application US/09463702A
; Patent No. 6335435
; GENERAL INFORMATION:
; APPLICANT: AGENE Research Institute, Co., Ltd.
; APPLICANT: HIRAKI AND ASSOCIATES
; APPLICANT: SHIMAMOTO AKIRO
; APPLICANT: KITNO, SAORI
; APPLICANT: FURUICHI, YASUHIRO
; TITLE OF INVENTION: HUMAN GENE RECO4 ENCODING HELICASE
; FILE REFERENCE: HIRAI150
; CURRENT APPLICATION NUMBER: US/09/463,702A
; PRIOR FILING DATE: 2000-01-24
; PRIOR APPLICATION NUMBER: PCT/JP98/03114
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: JAPAN 9/200387
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 37
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-463-702A-37

Query Match 29.3%; Score 1883; DB 3; Length 361;
Best Local Similarity 100.0%; Pred. No. 3.7e-144;
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 471 QALBOLGHOAFPPGGERAVMRILSGISTLLVPTGAGSLCYQLPALYSRRSPCLTIVV 530
DB 1 QALBOLGHOAFPPGGERAVMRILSGISTLLVPTGAGSLCYQLPALYSRRSPCLTIVV 60
QY 531 SPLSLMDQVSGLPPLCKAACISGMRKORESVLQKIRAAQVHVLMTPBALVGAGL 590
DB 61 SPLSLMDQVSGLPPLCKAACISGMRKORESVLQKIRAAQVHVLMTPBALVGAGL 120

QY 591 PPAOLPVAFACIDAEACLSQMSHNPFCYLKVCULREBVGVCFLGTATATRTAS 650
DB 121 PPAOLPVAFACIDAEACLSQMSHNPFCYLKVCULREBVGVCFLGTATATRTAS 180
QY 651 DVAOHLAAVEBBDLGPAPVPTNLHLSVSMDBDTQALITLLQGRFONLDSIIYYCNR 710
DB 181 DVAOHLAAVEBBDLGPAPVPTNLHLSVSMDBDTQALITLLQGRFONLDSIIYYCNR 240
QY 711 EDTERIALALRTCLHAANVPSGGAPKTTAAVHAAGCSRRRRVQAFMGOGLRVVVA 770
DB 241 EDTERIALALRTCLHAANVPSGGAPKTTAAVHAAGCSRRRRVQAFMGOGLRVVVA 300
QY 771 TVAFMGGLDRPDVRAVHLGLPSPFESYVQAVGRAGRGOPAHCHFLPOPGEDLRELR 830
DB 301 TVAFMGGLDRPDVRAVHLGLPSPFESYVQAVGRAGRGOPAHCHFLPOPGEDLRELR 360
QY 831 H 831
DB 361 H 361

RESULT 4
US-09-699-135-37
; Sequence 37, Application US/09699135
; Patent No. 6472513
; GENERAL INFORMATION:
; APPLICANT: AGENE Research Institute, Co., Ltd.
; APPLICANT: HIRAKI AND ASSOCIATES
; APPLICANT: SHIMAMOTO AKIRO
; APPLICANT: KITNO, SAORI
; APPLICANT: FURUICHI, YASUHIRO
; TITLE OF INVENTION: HUMAN GENS RECO4 ENCODING HELICASE
; FILE REFERENCE: HIRAI150
; CURRENT APPLICATION NUMBER: US/09/699,135
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US/09/463,702A
; PRIOR FILING DATE: 2000-01-24
; PRIOR APPLICATION NUMBER: PCT/JP98/03114
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: JAPAN 9/200387
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 37
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-699-135-37

Query Match 29.3%; Score 1883; DB 4; Length 361;
Best Local Similarity 100.0%; Pred. No. 3.7e-144;
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 471 QALBOLGHOAFPPGGERAVMRILSGISTLLVPTGAGSLCYQLPALYSRRSPCLTIVV 530
DB 1 QALBOLGHOAFPPGGERAVMRILSGISTLLVPTGAGSLCYQLPALYSRRSPCLTIVV 60
QY 531 SPLSLMDQVSGLPPLCKAACISGMRKORESVLQKIRAAQVHVLMTPBALVGAGL 590
DB 61 SPLSLMDQVSGLPPLCKAACISGMRKORESVLQKIRAAQVHVLMTPBALVGAGL 120
QY 591 PPAOLPVAFACIDAEACLSQMSHNPFCYLKVCULREBVGVCFLGTATATRTAS 650
DB 121 PPAOLPVAFACIDAEACLSQMSHNPFCYLKVCULREBVGVCFLGTATATRTAS 180
QY 651 DVAOHLAAVEBBDLGPAPVPTNLHLSVSMDBDTQALITLLQGRFONLDSIIYYCNR 710
DB 181 DVAOHLAAVEBBDLGPAPVPTNLHLSVSMDBDTQALITLLQGRFONLDSIIYYCNR 240
QY 711 EDTERIALALRTCLHAANVPSGGAPKTTAAVHAAGCSRRRRVQAFMGOGLRVVVA 770
DB 241 EDTERIALALRTCLHAANVPSGGAPKTTAAVHAAGCSRRRRVQAFMGOGLRVVVA 300

```

QY 771 TAAFGKGLDRPDVRAVHLGLPSPFESYQAVGRAGRDGQPAHCHLFLPOGEDLRELR 830
DB 301 TAAFGKGLDRPDVRAVHLGLPSPFESYQAVGRAGRDGQPAHCHLFLPOGEDLRELR 360
QY 831 H 831
DB 361 H 361

RESULT 5
US-08-781-891-75
; Sequence 75, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620tenburg, Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 607 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-781-891-75

Query Match 9.8%; Score 632.5; DB 3; Length 607;
Best Local Similarity 37.6%; Pred. No. 1.6e-42;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

QY 474 EQLGHAFFPGQGRAMRLISGISTLLVPLTGAGSLCYQLPALLYSRSPCLTLVVSPL 533
DB 18 ETFGVOQFRPGEBIIDTVLSGRDCLVWPTGGKSLCYQIPALLNG---LTVVVSPL 73
QY 534 ISLMDQVSGLRP-CLKAACTHSGMTRKQRESVLOKIRAAQVHVMLTPEALVGAGGLRP 592
DB 74 ISLMDQVQOLQANGVAAACLNSTGTREOQLEWMTGCRGQIRLTLTAPRLMDNFLH 133
QY 593 AAQLPVAACIDEAHCISQWSHNPFCYLVCKVLRBMGVHCFGLTATATRTASDV 652
DB 134 LHMNVVLLA-VDEAHCISQWGHDPFVYALGQ-LRGRFPLPLFMAITRATADTTRODI 191
QY 653 AQHLAAVEBDLHGPRVPVNTNHLVSMDRDTDOALLTLIGKRFQNDST-----703
DB 192 VRLIG-----LMDPL-----IOIS-SFDRPNIRVLM-----EKFRPLQLMKRYVOERG 235

```

```

QY 704 ---IYCNRRREDTERIALLTCTLHAAMVPGSGRAPKTTAEAYHAGMCSERRRRVORAF 760
DB 236 KSGITTCNSRAKVEETTAAL-----QSKGISAAAYAGLENNRADVQEKF 281
QY 761 MGGQLRVVATVAFGKGLDRPDVRAVHLGLPSPFESYQAVGRAGRDGQPAHCHLFL-- 818
DB 282 QRDLDQIVATVAFGKGLDRPDVRAVHLGLPSPFESYQAVGRAGRDGQPAHCHLFL-- 341
QY 819 -----QPGEDLRELRH---VHADSTDFLAVKRLV-----QVPPACTC- 855
DB 342 ADMAMLRRCLEEKPGQ-LQDIERHKLAMGAFAEAQTCRRLVLLNYGEGQEPGNCND 400
QY 856 TCTRPSEQEGAVGS 871
DB 401 ICLDPPKQYDGSTDAQ 416

RESULT 6
US-09-618-166-75
; Sequence 75, Application US/09618166
; Patent No. 6583112
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Intellectual Property Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/618,166
; FILING DATE: 17-Jul-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: McMaister, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 240052.419C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 607 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-618-166-75

Query Match 9.8%; Score 632.5; DB 4; Length 607;
Best Local Similarity 37.6%; Pred. No. 1.6e-42;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

QY 474 EQLGHAFFPGQGRAMRLISGISTLLVPLTGAGSLCYQLPALLYSRSPCLTLVVSPL 533
DB 18 ETFGVOQFRPGEBIIDTVLSGRDCLVWPTGGKSLCYQIPALLNG---LTVVVSPL 73
QY 534 ISLMDQVSGLRP-CLKAACTHSGMTRKQRESVLOKIRAAQVHVMLTPEALVGAGGLRP 592

```

```

Db 74 ISLMDQVDOQLANGVAAACLNSTOTREOQLEVMTCRTQIRLLIYIAPERLMLDNPLEH 133
QY 593 AAQLPVAACIDEAHCLISQSHNRPCLRVCKYLREMGVHCFGLTATATRTASDV 652
Db 134 LAHNPVLLA-VDEHCLISQWGHDFRPEYALGQ-LRQFPFLPFAALTATADTTRODI 191
QY 653 AQHLVAEPELDHGPAPVPTNLHLSVSMRDTDOALLTLQKRPONLDSI----- 703
Db 192 VRLLG-----LNDPL-----IQIS-SFRRPNIRYMLM-----EKFPDQLMRVYQEQRG 235
QY 704 ---IYCNRREDTERIALRLTCLHAANVPSSGGRAPKTTAAYAHGMSRRRRVQRAF 760
Db 236 KSGIITCNSRAKVEDTAAAL-----OSKGISAAAYHAGLENNVRADVQEKF 281
QY 761 MGGQLRVAVATVAFMGGLDRPDVRAVLHGLPPSFESYVQAVGRAGDQPAHCHLFL-- 818
Db 282 QRDQLQIVAVATVAFMGINKPVRVHFDPINIESYYQETGRAGDGLPAEAMLFYDP 341
QY 819 -----QPGEDLRELRH---VHADSTDFLAVKRLV-----QRFPACTG- 855
Db 342 ADMAWLRRCLEBKPGQG-LQDIERHKLAMGAPAEQOTRRLVLTNYFEGRGQEPGQND 400
QY 856 TCTRPPEQEGAVGE 871
Db 401 ICLDPPKQYDGSSTDAQ 416

```

RESULT 7
US-08-559-303B-77
Sequence 77, Application US/08559303B
Patent No. 5824501
GENERAL INFORMATION:
APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
APPLICANT: GRODEN
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
TITLE OF INVENTION: OF BLOOM'S SYNDROME
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/559,303B
FILING DATE: NOVEMBER 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/65
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 370
TYPE: AMINO ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: OTHER NUCLEIC ACID
HYPOTHEICAL: YES
ANTI-SENSE: NO
FEATURE:
NAME/KEY:
LOCATION:

```

; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; US-08-559-303B-77
Query Match 9.7%; Score 621; DB 2; Length 370;
Best Local Similarity 40.0%; Pred. No. 6,7e-42;
Matches 154; Conservative 57; Mismatches 110; Indels 64; Gaps 12;

```

```

QY 474 BGLGQARPPQGRARVMBILISITLVLPPGAGKSLCYOLPALLYSRRSPCLTVSPSL 533
Db 4 ETFGTYQFRPGQEEIITDVLSGRDCLVMPFGGKSLCYQIPALLING---LTVVSEPL 59
QY 534 ISLMDQVSGLPP-CLKAACIHSGMTKQRBSSVLQKIRPAQVHVMLTPREALVAGGLPP 592
Db 60 ISLMDQVDOQLANGVAAACLNSTOTREOQLEVMTCRTQIRLLIYIAPERLMLDNPLEH 119
QY 593 AAQLPVAACIDEAHCLISQSHNRPCLRVCKYLREMGVHCFGLTATATRTASDV 652
Db 120 LAHNPVLLA-VDEHCLISQWGHDFRPEYALGQ-LRQFPFLPFAALTATADTTRODI 177
QY 653 AQHLVAEPELDHGPAPVPTNLHLSVSMRDTDOALLTLQKRPONLDSI----- 703
Db 178 VRLLG-----LNDPL-----IQIS-SFRRPNIRYMLM-----EKFPDQLMRVYQEQRG 221
QY 704 ---IYCNRREDTERIALRLTCLHAANVPSSGGRAPKTTAAYAHGMSRRRRVQRAF 760
Db 222 KSGIITCNSRAKVEDTAAAL-----OSKGISAAAYHAGLENNVRADVQEKF 267
QY 761 MGGQLRVAVATVAFMGGLDRPDVRAVLHGLPPSFESYVQAVGRAGDQPAHCHLFL-- 818
Db 268 QRDQLQIVAVATVAFMGINKPVRVHFDPINIESYYQETGRAGDGLPAEAMLFYDP 327
QY 819 -----QPGEDLRELRH 831
Db 328 ADMAWLRRCLEBKPGQG-LQDIERH 351

```

RESULT 8
US-09-175-828-77
Sequence 77, Application US/09175828
Patent No. 6221643
GENERAL INFORMATION:
APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
APPLICANT: GRODEN
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
TITLE OF INVENTION: OF BLOOM'S SYNDROME
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/175,828
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/559,303
FILING DATE: NOVEMBER 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/65
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766


```

OY      AOHIAVAEBPDHGAAPVPTNHLTYSMDRTDQALTLGGRCPPNLSI-----703
Db      204 VRLG-----LNDEL-----IQVS-SFPRNIRYLM-----EKFPDQIMRYDDQGS 247
OY      704 ---11CNRREDTERIALLRITCLHAAWPGSGGARPKTTAEVYHAGMCSRRERRVORAE 760
Db      248 KSGIIIVCSNRSKVEEDTAAALQ-----SRG-----ISAAYAHGLENDVVAEVOEKE 293
OY      761 MGGOLRVVVAATYAFMGGLDRPDVPAVLHGLBPSFESYOAANGARBDGPAHCHFLQP 820
Db      294 QRDLDQIVVATYAFMGGLKPNVRFEVHNDIPRNIESTYQETGRAGRDGLPAPAMLFYDB 353
OY      821 -----QGEDLNELRRH---VHADSTDFLAFLRV-----QRVPACTG-T 856
Db      354 ADMAWLRRCLEBKRPAGPLODIERHKLINAMGAFAEAQTCRLVLYLNFEGEGRQSPGNCDI 413
OY      857 CTRPESBEGANGSGRPVK-YRQOEAELSH-----QAAPGRRCMCHERLPI 906
Db      414 CLDPKQDGLMDARALUSTYRVNDRFGMGVUEVYLRAGNNORIREMGIDK-LPV 468

```

```

RESULT 13
US-09-543-681A-4359
; Sequence 4359, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543, 681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 6344
; SEQ ID NO 4359
; LENGTH: 641
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4359

Query Match          9.2%; Score 591.5; DB 4; Length 641;
Best Local Similarity 31.9%; Pred. No. 3.7e-39;
Matches 187; Conservative 71; Mismatches 153; Indels 175; Gaps 25;

Oy 467 AEVFOAL-----EOLGHOAFPRGGERAWRIRISGTLIVYPTGAGKSLCYOLPML 518
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 36 AEVLNMSMSAOYLIRETFGTQOFRPGQEIITITIGRDLVMTPTGGKSLCYOIPML 95
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 519 YSRSPCLTVLVSPPLSLMDDOVSGLPCL--KAACIHSGMTRKQRESVLQKIRAAQVH 575
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 96 LDG---LTVVVSPLISLMDQVDDQL--CLHGIDAPLNFSTQSRDEQLQVWRCQREIK 149
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 576 VLMLTPRLVAGAGLPAAQLPVAFACIDEAHCLISQMSNHPRCYGLAVCKYLBERMGVH 635
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 150 LLVYAPESLMWESFHLHWQOP-ALLAVDEAHCHISQWGHFRPEY-IGIGILROYLPDV 207
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 636 CFLGATATARTASDVAAQHLVAAEEDLHGAPVPVTLHLSVSDMRDTDOALLTLGQK 695
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 PIALTAADMTTRDDIINQLNR-----TP--VHIS-SFDRNIRKTYLV---E 251
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 696 RFQNLDSI-----IYCNRREDTERIALALTCLHAAVPGSGGAPKTTAE 743
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 252 KYKPLDQLMLFIRGQKGKSGIITYCNSRSGKVBETAERL-----GKRG---LSIA 297
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 744 YHAGMCSERRRRVQRAFQGLRVVVAATVAGMGIDRPDVAVLHLGLPSPESFVQAVG 803
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 298 YHAGMDLATRAKVDAFQRPDLQIVATVAFGMGINKENVRFFVAFDIPRIEISYQETG 357
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 804 RAGRDGQPAHCHLFLQPGCEPLRELRRVHADSTDFLAVKRLVQGVFPACCTCTCRPSE 863
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 358 RAGRDGLPAQAVLFLVDP--ADVAMILR-----CLDEKPESE 391
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

[illegible]

```

RESULT 14
US-09-902-540-16497
: Sequence 16497, Application US/09902540
: Patent No. 6833447
: GENERAL INFORMATION:
: APPLICANT: Goldman, Barry S.
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: APPLICANT: Miegand, Roger C.
: TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
: FILE REFERENCE: 38-10115849)B
: CURRENT APPLICATION NUMBER: US/09/902, 540
: CURRENT FILING DATE: 2001-07-10
: PRIOR APPLICATION NUMBER: 60/217, 883
: PRIOR FILING DATE: 2000-07-10
: NUMBER OF SEQ ID NOS: 16825
: SEQ ID NO 16497
: LENGTH: 1394
: TYPE: PRT
: ORGANISM: Myxococcus xanthus
US-09-902-540-16497

Query Match      9.1%, Score 585; DB 4; Length 1394;
Best Local Similarity 26.8%; Pred. No. 3,9e-38;
Matches 261; Conservative 116; Mismatches 347; Indels 246; Gaps 44

QY      2  ERLADV--RERLQAW--ERAFR-----RGRGRPSQDDVEAAPEETRALYREYRLK 49
DB      240 QRYVDVLCRRLLQAHMEDHVMRVTTVITAVTSAGAPVAVDFQSTGTQVEGV--GMKVL 297
QY      50  RTTQAGGGLASSSESPLAAAEAP--EPRCMGPHLNRATKSPQPTPGSRGGSVVDYQ 107
DB      298 IGGQQAARLSEBGRKGDAERBEDEPQTLPSGLARQPOSVEDVEAVKKTTRPP--R 355
QY      108 RLKANKLKTLL--AGPAL-----GRRPWPLGRASSKA-----ST 139
DB      356 FTDAITLLTMMASGRALDEKELADMRGTGJTPTRAIIIVLLDREYLRGRKVMAT 415
QY      140 PK-----PPTGTPVSFAKVSDEDPQLPEPQPRFRLQHLQASLSQRGS 185
DB      416 EKGJHLIQVAVPDVKTPTMTQWEMWLQRIERGQQLDE-----FNGIEAYVLEVGH 469
QY      186 LDPGWLORCHSEVEDPDLAPKACRPDLG---SEESQLIPGESAVLPGAGSOGPEA--SA 241
DB      470 GATAPQRPFRPREV--LGAP-----PGAGWGCGGAQVASEGAAAHAGAGARGGCGGA 522
QY      242 FOEWSI--RVGSPQPSGSGEKKRRMNEBPWESPAPVQOESSQAQPPSEGAQAVAEEDPPG 300
DB      523 PAESSAERFGD---ASAGCAGRGCGAFTESA---ERFADAAPGVQCGR-----PG 569
QY      301 EPVQAQRPQPCSSSPNPRPHGLSPSSQARAKAEBTALHIFPRLARHDRGNVYLVLMKQ 360
DB      570 ALSESESD-----LFGRRKRTGSGVSRSEPH-----GEPPLSTMSR 606
QY      361 KHYVGRALRSRLRKQ--AWKQ--KWRKKGEFGGAGATVTTKSCFLNEQPDHMAOCP 417
DB      607 -----CGGAARLRLPRADGTGMNAGAMENTG-----AARAP 636
QY      418 -----RPASEEDTDAVGEPLVPSQVPEVPSLDPTVLPLYSLSQPSQALATPRAEVQ 472

```

```

Db      637 VNRARTPASAATG-----QRRERVARAPTAPALRFLK-----671
Qy      473 LEOLGHOAFPCQOEAWRILSGISTLVLPFGAGKSLCYOLPALIYSRSPCLTLVSP 532
Db      672 -EAFSPDFRPOEAVCRATAGBEDLLVMPGAGKSLCYOLPGLARAG---TTLVSP 726
Qy      533 LLSLMDDOVSGLPK-LKACIHSQWTRKQRESVLQKIRAAQVHVMLEALPVAGGSP 591
Db      727 LIALMEDOVARLOSIGFADRIHSGDRMSROVCADYIEDRLDFLFIAPERLGVPGFYE 786
Qy      592 PAQOLPVAACIDEAHCLISQWSHNPFCYLKRVCKLREMGV---HCFIQTATATART 648
Db      787 FLARATP-ALIAVDEHCHISQWGHDRPXY-----RLGARLPULRARPVALTATATPRV 841
Qy      649 ASDVAQHLAABEEDLHGAPVP-----TNLHLSVSMDDTD---QALLTLQ 694
Db      842 QRDIVQOLG-----LQPGGKARFTFHGRRTNIAIEV---RELNPGARGDAIQGLEED 892
Qy      695 KRFOULDSIIICNREDTERIAALLRTCLHAAWPGSGGAPKTTAEVHAGMCSRRER 754
Db      893 ENR---PAIVAAATRKABEQLADOL-----AGEPP---AAAYHAGLOPSEED 934
Qy      755 RVQAFPMQGLRVVAVTAFGMLDRPDVRAVLHLGLPPSPESYQAVAGRAGDQPAHC 814
Db      935 RVQAFELGSLLEVIAITTFGMDKADVRVIAHALPASLBYQELGRAGRDGKPSRA 994
Qy      815 ---HLFLOPQEDLRELRRHVADSTDFLAVKRLVQVFPACTCTTRPPSBOGAVGE 871
Db      995 VLLHSYIDRRTHEFFRRDYPEA-----YLERLFKS-TAPQLEBKAVLQGRVGD 1044
Qy      872 RVPKYPPOEAEOL 885
Db      1045 ---PEVFDALAEOL 1055

RESULT 15
US-09-134-000C-5710
; Sequence 5710, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patent version 3.1
; SEQ ID NO 5710
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5710

```

Query Match 8.6%; Score 554.5; DB 4; Length 589;
 Best Local Similarity 34.0%; Pred. No. 3,2e-36;
 Matches 143; Conservative 67; Mismatches 147; Indels 63; Gaps 10;

```

Qy      472 ALEOL-----GHQAFRQGEARVAMRILSGISTLVLPFGAGKSLCYOLPALIYSRSPCL 526
Db      3 ALQELLLKDTFGYDDRFPGQETIIRVLRQENVLGIMPTGGKSLCYOLPALLLDN---L 58
Qy      527 TLVSPPLSLMDQVS-----GLPPLKACIHSQWTRKQRESVLQKIRAAQVHVMLEP 581
Db      59 TLVISPLSLMDQVDALNMGIP-----ATYINSTISYQEMNHRITOLANKEVKLLYAP 114
Qy      582 EALVAGGLPPAQOLPVAACIDEAHCLISQWSHNPFCYLKRVCKLREMGVHCFGLIT 641
Db      115 ERLSESDFQOQMLTHV-PIDLLAVDEAHCLISQWGHDRFSPYRLAEITIDQFOOQPTVIALT 173

```

```

Qy      642 ATATRTASDVNOHLAABEEDLHGAPVPVNTLHLSVSMDDRTDQALLTLQKRFONLD 701
Db      174 ATATPOVAEDIVKOLRIPSENEIK-TGFARENLSQVYKQNDVFLFIYLNMTGQ---229
Qy      702 SIIICNREDTERIAALLRTCLHAAWPGSGGAPKTTAEVHAGMCSRRERRVQRAFEM 761
Db      230 SGIIVASTRKEVERIYHLL-----ESKTLAAGMYHGMSPQLRSENQEAFL 275
Qy      762 QGOLRVVAVTAFGMLDRPDVRAVLHLGLPPSPESYQAVAGRAGDQPAHCFLFLOPQ 821
Db      276 YDQVQVMVATNAFGMINKSNVRFVIAQVPGNIESYQOEAAGRAGDGLPSDAVLMFAPO 335
Qy      822 GED-----LRELRRHVADSTDFLAVKRLVQVFPACTCTTR 859
Db      336 DLQIQYFLBQSEMTIDYKQKEYLKLREMSQYANAO---MCLQYKILRYFGBEGTDGCR 391

```

Search completed: March 19, 2005, 01:08:46
 Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2005, 00:54:10 ; Search time 180 Seconds
(without alignment)

2595.592 Million cell updates/sec

Title: US-09-889-325-4

Perfect score: 6424

Sequence: 1 MERLRDVRERQAMERARR.....HLSTHALVGLATEELLQVAR 1208

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: geneseqp1980s:*

2: geneseqp1990s:*

3: geneseqp2000s:*

4: geneseqp2001s:*

5: geneseqp2002s:*

6: geneseqp2003as:*

7: geneseqp2003bs:*

8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6424	100.0	1208	3	AA020993 Human Rec
2	6424	100.0	1208	5	ABG93387 Human Rec
3	6424	100.0	1208	8	ADQ21562 Human Rec
4	6421	100.0	1208	2	AAW95050 Human Rec
5	6085.5	94.7	1306	8	ADM90956 Human Rec
6	6050	94.2	1144	7	ADP01202 Human Rec
7	6029	93.9	1142	8	ABM84821 Human Rec
8	4006	62.4	756	3	AA020995 Human Rec
9	2820	45.5	554	3	AA020994 Human Rec
10	1723.5	26.8	1579	4	ABG63317 Drosophila
11	631.5	9.8	610	4	AAU34821 E. coli
12	631.5	9.8	610	4	ABU15433 Protein e
13	619	9.6	609	6	ABU28286 Protein e
14	609.5	9.5	609	6	ABU47859 Protein e
15	609.5	9.5	615	4	AAU38154 Salmonella
16	607.5	9.5	615	4	ABU47049 Protein e
17	606	9.4	608	6	ABU32005 Protein e
18	606	9.4	619	7	ABO61558 Klebsiella
19	593	9.2	610	7	ABU50101 Protein e
20	593.5	9.2	641	7	ADP04074 Bacteriophage
21	589.5	9.2	600	6	ABU41290 Protein e
22	586.5	9.1	613	6	ABM68051 Photorehab
23	581.5	9.0	620	6	ABU49054 Protein e
24	578.5	9.0	632	6	ABU39369 Protein e
25	578	9.0	619	4	AAU35492 Haemophil

ALIGNMENTS

26	578	9.0	619	6	ABU30347 Protein e
27	554.5	8.6	589	7	ABU14523 Protein e
28	554.5	8.6	589	7	ADH87825 Enterococ
29	552	8.6	590	5	ABD48903 Libertia
30	552	8.6	590	6	ABU32767 Protein e
31	549	8.5	597	6	ABU44940 Protein e
32	547.5	8.5	589	4	AAU35210 Enterococ
33	547	8.5	589	7	ABO69207 Pseudomon
34	545	8.5	580	4	AAU33463 Enterococ
35	545	8.5	712	6	AAU36322 Pseudomon
36	545	8.5	712	6	ABU38552 Protein e
37	544.5	8.5	390	6	ABU33517 Protein e
38	544	8.5	341	8	ADN26337 Bacteriophage
39	542.5	8.4	1955	3	AAU94670 Bacteriophage
40	536	8.3	651	5	ABP65837 Bacteriophage
41	529	8.2	616	7	ADP97631 Bacteriophage
42	528	8.2	590	6	ABU29694 Protein e
43	526	8.2	592	6	ABU42752 Protein e
44	526	8.2	637	5	ABP40813 Staphylococ
45	526	8.2	637	8	ADP08152 Staphylococ

RESULT 1

AA020993 standard; protein; 1208 AA.

AA020993;

11-DEC-2000 (first entry)

Human RecQ4 helicase.

RecQ4 helicase; human; Rothmund-Thomson syndrome; chromosome 8q24.3;

poikiloderma congenitale; autosomal recessive; skin disorder;

dermatology; antibody; prenatal diagnosis; gene therapy.

Homo sapiens.

WO200043522-A1.

27-JUL-2000.

19-JAN-2000; 2000WO-JP000233.

19-JAN-1999; 99JP-00011218.

(AGEN-) AGENE RES INST CO LTD.

Kitao S, Shimamoto A, Furuichi Y;

WPI; 2000-524241/47.

N-PSDB; AAA72320, AAA72321.

RecQ4 helicase gene, gene products and antibody, used in the diagnosis

and treatment of Rothmund-Thomson syndrome, e.g. by gene therapy.

Example 5; Page 83-92; 115pp; Japanese.

The present sequence represents human RecQ4 helicase. The invention relates to the genomic DNA sequence of human RecQ4 helicase (AA020993). Mutations in the RecQ4 helicase gene, located on chromosome 8q24.3, are the cause of Rothmund-Thomson syndrome (also known as poikiloderma congenitale), an autosomal recessive skin disorder principally occurring in females and often accompanied by juvenile cataracts, saddle nose, congenital bone defects, hypogonadism and disturbances in the growth of hair, nails and teeth. The invention also relates to vectors and host cells comprising the human RecQ4 helicase genomic sequence. It additionally encompasses use of the RecQ4 helicase protein as a therapeutic and anti-RecQ4 antibodies as diagnostic agents. The RecQ4 helicase gene and its products, and anti-RecQ4 helicase antibodies are

CC useful in the diagnosis, especially prenatal diagnosis, and treatment of
 CC Rothmund-Thomson syndrome. The genomic sequence may especially be used in
 CC gene therapy for this condition

xx Sequence 1208 AA;

Query Match 100.0%; Score 6424; DB 3; Length 1208;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MERLADVERLQAMRRARRRQGRPPSODDVEAAEEETRALYREYRTLKTGGAGGGR 60
DB 1 MERLADVERLQAMRRARRRQGRPPSODDVEAAEEETRALYREYRTLKTGGAGGGR 60
QY SSESIPAAAEBAERBCRGPHLNRAATKSPOPTPKRSQGSVPDQGORLKANLKGTLQAG 120
DB SSESIPAAAEBAERBCRGPHLNRAATKSPOPTPKRSQGSVPDQGORLKANLKGTLQAG 120
QY 121 PALGRPMPPLGRASSKASTPKPGTGPVPSFAKVSDEPPQLPEPQPRGRLOHLQASIS 180
DB 121 PALGRPMPPLGRASSKASTPKPGTGPVPSFAKVSDEPPQLPEPQPRGRLOHLQASIS 180
QY 181 QRLGSLDGMQORCHSEVPDFLGAPKACRPDLGSEBSQLLTGESAVLGPAGSGSPEAS 240
DB 181 QRLGSLDGMQORCHSEVPDFLGAPKACRPDLGSEBSQLLTGESAVLGPAGSGSPEAS 240
QY 241 AFOEVSIVGSPQSSSGGKKRNEBPWESPAPQVQOESSQAGPSEGAVAVEDPPG 300
DB 241 AFOEVSIVGSPQSSSGGKKRNEBPWESPAPQVQOESSQAGPSEGAVAVEDPPG 300
QY 301 EPVQAPQPPQSSSPNPRYHGLSPSSQARAGAEGTAPLHIFRLARDRGNYRLNKKQ 360
DB 301 EPVQAPQPPQSSSPNPRYHGLSPSSQARAGAEGTAPLHIFRLARDRGNYRLNKKQ 360
QY 361 KHYVGRALRSRLAKQAMQKRRKGCFCGGGATVTTKESCFINEQPDHAAACPRPA 420
DB 361 KHYVGRALRSRLAKQAMQKRRKGCFCGGGATVTTKESCFINEQPDHAAACPRPA 420
QY 421 SEEDTDAGPEPLVSPQVPEVPSLDPTVLPLYSLSGSGQAEPAEPAVFOLEQGHQA 480
DB 421 SEEDTDAGPEPLVSPQVPEVPSLDPTVLPLYSLSGSGQAEPAEPAVFOLEQGHQA 480
QY 481 FRPQGERAVMRILSGISTLLVPTGAGKSLCYQLPALYSRRSPCLTVVSPLSLMDQ 540
DB 481 FRPQGERAVMRILSGISTLLVPTGAGKSLCYQLPALYSRRSPCLTVVSPLSLMDQ 540
QY 541 VSGLRPCLKAACIHSGMTRKORESVLQKIRAAQVHMLTPBALVAGGLPPAQLPVA 600
DB 541 VSGLRPCLKAACIHSGMTRKORESVLQKIRAAQVHMLTPBALVAGGLPPAQLPVA 600
QY 601 FACIDEAHLCSQMSHNPFCYLRVCKVLERMGVHCFLGTATATRTASDVAOHLAAVE 660
DB 601 FACIDEAHLCSQMSHNPFCYLRVCKVLERMGVHCFLGTATATRTASDVAOHLAAVE 660
QY 661 EPDHGPAPVPTNLHLSVMDRDTQALLTLLOGRFQNDLSIIYCNREDTERIALAL 720
DB 661 EPDHGPAPVPTNLHLSVMDRDTQALLTLLOGRFQNDLSIIYCNREDTERIALAL 720
QY 721 RTCLHAANVPSSGGGAPKTTAAATAGMCSRERRRVOAFMGGQRLVVATATAFGMDR 780
DB 721 RTCLHAANVPSSGGGAPKTTAAATAGMCSRERRRVOAFMGGQRLVVATATAFGMDR 780
QY 781 PDVRAVHLGLPPSPESYVQAVGRAGRDQPAHCHFLQPOEEDRELRRHHAADSTPL 840
DB 781 PDVRAVHLGLPPSPESYVQAVGRAGRDQPAHCHFLQPOEEDRELRRHHAADSTPL 840
QY 841 AVKRLVQRFPACTCTCTRPSEOGAVGGERPVKYPQOABOJSHQAAPRRVCMGH 900
DB 841 AVKRLVQRFPACTCTCTRPSEOGAVGGERPVKYPQOABOJSHQAAPRRVCMGH 900
QY 901 BRALPIQLTVQALDMPBEBAIEITLICYLBHPHHMLBLATTTTHRLNCGGPAQLQALA 960
DB 901 BRALPIQLTVQALDMPBEBAIEITLICYLBHPHHMLBLATTTTHRLNCGGPAQLQALA 960

```

```

QY 961 HRCPELAVCLAQQLPEDEGQSSSVFPMVXLVDSMGWELASVRRALCOLQMDHEPRITGV 1020
DB 961 HRCPELAVCLAQQLPEDEGQSSSVFPMVXLVDSMGWELASVRRALCOLQMDHEPRITGV 1020
QY 1021 RRGTVLVEFSELAFHLSPGDLTAEEKQICDPLFYGVQVQARERQALRLRRTFOAFHSV 1080
DB 1021 RRGTVLVEFSELAFHLSPGDLTAEEKQICDPLFYGVQVQARERQALRLRRTFOAFHSV 1080
QY 1081 APPSGPCLQEOODERSTRLDLGRYEEBEGSGPGMEDAQGPQOARLODEDDVR 1140
DB 1081 APPSGPCLQEOODERSTRLDLGRYEEBEGSGPGMEDAQGPQOARLODEDDVR 1140
QY 1141 CDIRQFLSRPEEKSSRAVARIFHGIGSPCYPAQVYQODRRFWRKYLHLSFHALVGLAT 1200
DB 1141 CDIRQFLSRPEEKSSRAVARIFHGIGSPCYPAQVYQODRRFWRKYLHLSFHALVGLAT 1200
QY 1201 EELLQVAR 1208
DB 1201 EELLQVAR 1208

```

```

RESULT 2
ABG93387
ID ABG93387 standard; protein; 1208 AA.
XX
AC ABG93387;
XX
DT 20-NOV-2002 (first entry)
XX
DE Human RecQ protein-like 4 (RECQL4).
XX
KW Human; RecQ protein-like 4; RECQL4; infection; inflammation; cytostatic;
KW tumour formation; cancer; antiinflammatory; antimicrobial;
KW antisense therapy.
XX
OS Homo sapiens.
XX
PN US6436706-B1.
XX
PD 20-AUG-2002.
XX
PF 23-FEB-2001; 2001US-00792594.
XX
PR 23-FEB-2001; 2001US-00792594.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ward DT, Watt AT;
XX
DR WPI; 2002-689941/74.
XX
N-PSDB; ABS68851.
XX
PT New antisense compounds targeted to nucleic acids encoding RecQ protein-
PT like 4, useful for modulating expression of the nucleic acid and treating
PT diseases associated with expression of the nucleic acid in humans.
XX
PS Disclosure: Col 47-56; 45pp; English.
XX
CC The invention relates to a compound targeted to specific nucleobases of
CC RecQ protein-like 4 (RECQL4) and which hybridises and inhibits the
CC expression of RECQL4. The compound is useful for inhibiting the
CC expression of RECQL4 in cells or tissues and for treating an animal,
CC particularly a human suspected of having or being prone to a disease or
CC condition associated with expression of RECQL4. The compound is useful
CC for diagnostics, therapeutics and as a research reagent, e.g.
CC prophylactically to prevent or delay infection, inflammation or tumour
CC formation. This sequence represents human RECQL4 polypeptide
XX
SQ Sequence 1208 AA;

```

Query Match 100.0%; Score 6424; DB 5; Length 1208;
 Best Local Similarity 100.0%; Pred. No. 0;

Matches	1208:	Conservative	0:	Mismatches	0:	Indels	0:	Gaps	0:
Qy	1	MERLADVREIRLOAMERARRRGRPPSODDVEAAPEETRALYREYRTTKRTTGAAGGGLR	60						
Dd	1	MERLADVREIRLOAMERARRRGRPPSODDVEAAPEETRALYREYRTTKRTTGAAGGGLR	60						
Qy	61	SESELPAAAEAEPERCGPHLNRAATKSPPTPKRSQGSVPDQORLKAUKLTQLOAG	120						
Dd	61	SESELPAAAEAEPERCGPHLNRAATKSPPTPKRSQGSVPDQORLKAUKLTQLOAG	120						
Qy	121	PALGRPMPPLGRASSKASTPKPGTGPVPSFAEKVSDEBPQLPEBPGRGLQHLQASLS	180						
Dd	121	PALGRPMPPLGRASSKASTPKPGTGPVPSFAEKVSDEBPQLPEBPGRGLQHLQASLS	180						
Qy	181	QRLGSLDGMLORCHESEVPDLGAPKACRPDLGSEBSQULLPGBSAVLGPAGSGSPRAS	240						
Dd	181	QRLGSLDGMLORCHESEVPDLGAPKACRPDLGSEBSQULLPGBSAVLGPAGSGSPRAS	240						
Qy	241	AFOEVSIVGSPOPSSSGEKKRKNBEPWESPAPVOQSSSQAGPSPSEGAVAVEEDPPG	300						
Dd	241	AFOEVSIVGSPOPSSSGEKKRKNBEPWESPAPVOQSSSQAGPSPSEGAVAVEEDPPG	300						
Qy	301	EPVOAQPPOCSSSPNPRYHGLSPSSQARAGKAGETAPLHIFPRLARHDGRNYVRLANKQ	360						
Dd	301	EPVOAQPPOCSSSPNPRYHGLSPSSQARAGKAGETAPLHIFPRLARHDGRNYVRLANKQ	360						
Qy	361	KHYVGRALRSRLRKQAMKQKMKKGGCGGATVTTKESCFINBOPDHMAOCCPRPA	420						
Dd	361	KHYVGRALRSRLRKQAMKQKMKKGGCGGATVTTKESCFINBOPDHMAOCCPRPA	420						
Qy	421	SEEDDAVGPEPLVSPQVPEVPSLDPTVLPLYSLGPSCQALAEPAEFQALBOLGHOA	480						
Dd	421	SEEDDAVGPEPLVSPQVPEVPSLDPTVLPLYSLGPSCQALAEPAEFQALBOLGHOA	480						
Qy	481	FRPGEBAVMBILSGISSTLVLP.TGAGKSLCYQLPALYSRRSPCLTLVSPSLSLMDQ	540						
Dd	481	FRPGEBAVMBILSGISSTLVLP.PTGAGKSLCYQLPALYSRRSPCLTLVSPSLSLMDQ	540						
Qy	541	VSGLPCCIAKACIHSGMTRKQRESYLOKIRAAQVHMLTPPALVAGAGLPPAAOLPVA	600						
Dd	541	VSGLPCCIAKACIHSGMTRKQRESYLOKIRAAQVHMLTPPALVAGAGLPPAAOLPVA	600						
Qy	601	FACIDEAHCLSMWSNFRPCYLAVGVKLBEMGVHCPGLTATARTASDVQAOLAAVAB	660						
Dd	601	FACIDEAHCLSMWSNFRPCYLAVGVKLBEMGVHCPGLTATARTASDVQAOLAAVAB	660						
Qy	661	EPDLHGPAPVPTNLHL.SVSMRDITDQALLTLQGRFQNLDSIIICYNRREDTERIALTL	720						
Dd	661	EPDLHGPAPVPTNLHL.SVSMRDITDQALLTLQGRFQNLDSIIICYNRREDTERIALTL	720						
Qy	721	RTCLHAANVPSSGGAPRTTAAVYAHGMCSSRRRRVQRAFMQGLRVVATYAFQMGIDR	780						
Dd	721	RTCLHAANVPSSGGAPRTTAAVYAHGMCSSRRRRVQRAFMQGLRVVATYAFQMGIDR	780						
Qy	781	PDVRAVLHGLPPSESVVQAVGRAGRGQPAHCHLPLQPOSEDLRELRRHNAHSTDFL	840						
Dd	781	PDVRAVLHGLPPSESVVQAVGRAGRGQPAHCHLPLQPOSEDLRELRRHNAHSTDFL	840						
Qy	841	AVKRLVQVFPACTCTCTRPSEBOGAVGGERPVFKYPOBAEQLSHOQAAPRRVCMGH	900						
Dd	841	AVKRLVQVFPACTCTCTRPSEBOGAVGGERPVFKYPOBAEQLSHOQAAPRRVCMGH	900						
Qy	901	ERALPIQLTVQALDMPBEAIEITLLCYLSLHPHMLLELATTYTHRLNCPGSPALQALA	960						
Dd	901	ERALPIQLTVQALDMPBEAIEITLLCYLSLHPHMLLELATTYTHRLNCPGSPALQALA	960						
Qy	961	HRCPLAVCLAQQLBEDPGSSSVYPMVVKLVDSMGWELASVRALCOLQMDHDEPRGV	1020						
Dd	961	HRCPLAVCLAQQLBEDPGSSSVYPMVVKLVDSMGWELASVRALCOLQMDHDEPRGV	1020						
Qy	1021	RRTGTVLVEFSELAFLHRSFGDLTAEBKQICDPLGYRQVQARERQALRLRTFOAFHSV	1080						
Dd	1021	RRTGTVLVEFSELAFLHRSFGDLTAEBKQICDPLGYRQVQARERQALRLRTFOAFHSV	1080						

Qy	1081	APPSGCLTEQODEBSRSLKDLGRVFEESGQEPGGMEDAAQGPESQARLQDMEDQVR	1140						
Dd	1081	APPSGCLTEQODEBSRSLKDLGRVFEESGQEPGGMEDAAQGPESQARLQDMEDQVR	1140						
Qy	1141	CDIRQFLSLRPEEKFSRAVARIFHIGSPCYPAQVYGQDRFMRKYLHLSFHALVGLAT	1200						
Dd	1141	CDIRQFLSLRPEEKFSRAVARIFHIGSPCYPAQVYGQDRFMRKYLHLSFHALVGLAT	1200						
Qy	1201	ELLQVAR 1208							
Dd	1201	ELLQVAR 1208							
RESULT 3									
ID	ADQ21562	standard; protein; 1208 AA.							
AC	ADQ21562;								
DT	26-AUG-2004	(first entry)							
DE	Human soft tissue sarcoma-upregulated protein - SEQ ID 4382.								
KW	soft tissue sarcoma; cytosstatic; gene therapy; vaccine; screening; human.								
OS	Homo sapiens.								
PN	WO2004048938-A2.								
PD	10-JUN-2004.								
PF	26-NOV-2003; 2003WO-US038193.								
PR	26-NOV-2002; 2002US-0429739P.								
PA	(PROT-) PROTEIN DESIGN LABS INC.								
PI	Aziz N, Ginsburg WM, Zlotnick A;								
DR	WPI; 2004-441208/41.								
PT	Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.								
PS	Example 2; SEQ ID NO 4382; 210bp; English.								
CC	The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytosstatic applications and may be useful for detecting soft tissue sarcoma, possibly via gene therapy or vaccine production. The nucleic acid sequences may be useful in diagnostic and screening applications. The current sequence is that of a human soft tissue sarcoma-upregulated protein of the invention. The current sequence is not shown within the specification per se but was submitted in CD format by the inventor.								
XX	Sequence 1208 AA;								
Query Match									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches	1208:	Conservative	0:	Mismatches	0:	Indels	0:	Gaps	0:
Qy	1	MERLADVREIRLOAMERARRRGRPPSODDVEAAPEETRALYREYRTTKRTTGAAGGGLR	60						
Dd	1	MERLADVREIRLOAMERARRRGRPPSODDVEAAPEETRALYREYRTTKRTTGAAGGGLR	60						

```
QY 61 SSESIPAAAEBAPEBRCWCPHLNRAATKSPPTPGRSQSVDPYQGRUKANLKGTLQAG 120
Db 61 SSESIPAAAEBAPEBRCWCPHLNRAATKSPPTPGRSQSVDPYQGRUKANLKGTLQAG 120
QY 121 PALGRPMPPLGRASSKASTPKRPGTGPVPSFAEKVSDEPPOLPEQPRGRLOHLQASLS 180
Db 121 PALGRPMPPLGRASSKASTPKRPGTGPVPSFAEKVSDEPPOLPEQPRGRLOHLQASLS 180
QY 181 QRLGSLDGMWLORCHSEVPDFLGAPKACRPDLGSEESQLLIPGESAVLGPAGSQPEAS 240
Db 181 QRLGSLDGMWLORCHSEVPDFLGAPKACRPDLGSEESQLLIPGESAVLGPAGSQPEAS 240
QY 241 AFQEVSIIVGSPQPSSSGGEKRRNNEEPWESPDAVQOESSQAGPPSEGAAGAVAEEDPPG 300
Db 241 AFQEVSIIVGSPQPSSSGGEKRRNNEEPWESPDAVQOESSQAGPPSEGAAGAVAEEDPPG 300
QY 301 EPVQAPQPPQCSSPNNPRYHGLSPSSQARAGKAEETAPLHLPRLARHDRGNVRLNKKQ 360
Db 301 EPVQAPQPPQCSSPNNPRYHGLSPSSQARAGKAEETAPLHLPRLARHDRGNVRLNKKQ 360
QY 361 KHYVGRALRSRLRKQAMKQKMRKKGECFGGAGATVTTKESCFINEQFDHMAAQCPRPA 420
Db 361 KHYVGRALRSRLRKQAMKQKMRKKGECFGGAGATVTTKESCFINEQFDHMAAQCPRPA 420
QY 421 SEEDTAVGPEPLVSPQVPPEVPSLDPTVLPLYSLSGSGQLAETPAEVFQALEQLGHQA 480
Db 421 SEEDTAVGPEPLVSPQVPPEVPSLDPTVLPLYSLSGSGQLAETPAEVFQALEQLGHQA 480
QY 481 FRPQGERAVMRLISGISLTLVLPFGAGKSLCYQLPALLYSRSRSLTLVVSGLSLIMDQ 540
Db 481 FRPQGERAVMRLISGISLTLVLPFGAGKSLCYQLPALLYSRSRSLTLVVSGLSLIMDQ 540
QY 541 VSGLPPLCAACIHSGMTKQRESVLOKIRAAQVHVMLTPEALVAGAGLPPAOLPVA 600
Db 541 VSGLPPLCAACIHSGMTKQRESVLOKIRAAQVHVMLTPEALVAGAGLPPAOLPVA 600
QY 601 FACIDEAHCLSQWSHNPFCYLKVCVLRERMGVHCFLGLTATATRTASDVAQHLAAVE 660
Db 601 FACIDEAHCLSQWSHNPFCYLKVCVLRERMGVHCFLGLTATATRTASDVAQHLAAVE 660
QY 661 EPDLHGAPVPTNMLLSVSMRBDTQALLTLQGRFQNLDSIIYYCNRRRETERIAALL 720
Db 661 EPDLHGAPVPTNMLLSVSMRBDTQALLTLQGRFQNLDSIIYYCNRRRETERIAALL 720
QY 721 RTCLHAANVPSSGSGRAPKTAAEAHAAGMCSRERRRQVQAFMGOQLRVVATAFGMDR 780
Db 721 RTCLHAANVPSSGSGRAPKTAAEAHAAGMCSRERRRQVQAFMGOQLRVVATAFGMDR 780
QY 781 PDVRAVHLGLPSPFESYVQAVGARGDQPAHCHLPLQPGEDLRELRHVHADSTDFL 840
Db 781 PDVRAVHLGLPSPFESYVQAVGARGDQPAHCHLPLQPGEDLRELRHVHADSTDFL 840
QY 841 AVKRLVQVFPACTCTCTRPSPSEOGAVGGERPVPRKPYQGEBOJSHQAAPEPRVCMGH 900
Db 841 AVKRLVQVFPACTCTCTRPSPSEOGAVGGERPVPRKPYQGEBOJSHQAAPEPRVCMGH 900
QY 901 ERALPIQTLVQALMDPEBAIEFLCYLELHPHMLIELATTYTHCRINCPGPAOLOALA 960
Db 901 ERALPIQTLVQALMDPEBAIEFLCYLELHPHMLIELATTYTHCRINCPGPAOLOALA 960
QY 961 HRCPLAVLAQQLPEDPGQSSSVFPMVKLVDSMGWELASVRBALCOLQMDHBPRTGV 1020
Db 961 HRCPLAVLAQQLPEDPGQSSSVFPMVKLVDSMGWELASVRBALCOLQMDHBPRTGV 1020
QY 1021 RRGTVAVFESFLAHLSPGDLTAEBKQICDELIXGVQAEERQALRLRTTQAFISV 1080
Db 1021 RRGTVAVFESFLAHLSPGDLTAEBKQICDELIXGVQAEERQALRLRTTQAFISV 1080
QY 1081 AFPSGCPCLQOQDERSTRKDLGRVFESEGEQPGMEDAQGEPPQARLQDMEDQVR 1140
Db 1081 AFPSGCPCLQOQDERSTRKDLGRVFESEGEQPGMEDAQGEPPQARLQDMEDQVR 1140
QY 1141 CDIROPFLSLRBEKESSRAVARIIPHIGSPCYPAQVYGQDRRFWRKYLHLSFHALVGLAT 1200
```

```
Db 1141 CDIROPFLSLRBEKESSRAVARIIPHIGSPCYPAQVYGQDRRFWRKYLHLSFHALVGLAT 1200
QY 1201 BELLOVAR 1208
Db 1201 BELLOVAR 1208

RESULT 4
AAM95050
ID AAM95050 standard; protein; 1208 AA.
XX
XX AAM95050;
AC
XX
XX 14-MAY-1999 (first entry)
DT
XX
DE Human helicase protein.
XX
XX RecQ4 gene; helicase; Werner's syndrome; Bloom's syndrome; human.
XX
XX Homo sapiens.
XX
XX W09905284-A1.
XX
XX 04-FEB-1999.
PD
XX
XX 10-JUL-1998; 98WO-JP003114.
PF
XX
XX 25-JUL-1997; 97JP-00200387.
PR
XX
XX (AGEN-) AGENE RES INST CO LTD.
PA
XX
XX Shitamoto A, Kitao S, Furutachi Y;
PI
XX
XX MPI; 1999-142939/12.
DR
XX
XX N-PSDB; AAK21656.
PT
XX
XX New human helicase gene RecQ4 - used for investigation and diagnosis of
PR helicase-implicated diseases such as Werner's syndrome.
XX
XX
XX Claim 1; Page 35-42; 67bp; Japanese.
PS
XX
XX The present sequence represents a protein having helicase activity
CC encoded by the human gene RecQ4. The gene has significant homology to the
CC Escherichia coli helicase gene (RecQ). Host cells transformed with
CC vectors comprising the RecQ4 gene are used for the recombinant expression
CC of the protein. The gene may be used for the study and diagnosis of
CC disorders in which helicase activity is involved, such as Werner's and
CC Bloom's syndromes in which mutations in the helicase gene are implicated
XX
XX
XX Sequence 1208 AA;

Query Match 100.0%; Score 6421; DB 2; Length 1208;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1207; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MERLADVVERLOQAMERAFRRORGRPSODDVEAAPEETRALYREYRTLKRTTGAAGGLR 60
Db 1 MERLADVVERLOQAMERAFRRORGRPSODDVEAAPEETRALYREYRTLKRTTGAAGGLR 60
QY 61 SSESIPAAAEBAPEBRCWCPHLNRAATKSPPTPGRSQSVDPYQGRUKANLKGTLQAG 120
Db 61 SSESIPAAAEBAPEBRCWCPHLNRAATKSPPTPGRSQSVDPYQGRUKANLKGTLQAG 120
QY 121 PALGRPMPPLGRASSKASTPKRPGTGPVPSFAEKVSDEPPOLPEQPRGRLOHLQASLS 180
Db 121 PALGRPMPPLGRASSKASTPKRPGTGPVPSFAEKVSDEPPOLPEQPRGRLOHLQASLS 180
QY 181 QRLGSLDGMWLORCHSEVPDFLGAPKACRPDLGSEESQLLIPGESAVLGPAGSQPEAS 240
Db 181 QRLGSLDGMWLORCHSEVPDFLGAPKACRPDLGSEESQLLIPGESAVLGPAGSQPEAS 240
QY 241 AFQEVSIIVGSPQPSSSGGEKRRNNEEPWESPDAVQOESSQAGPPSEGAAGAVAEEDPPG 300
```

Db 241 AFGVSIKRGSPQSSGGEKRMNEPESPAQVQESSQAGPSEBGAGAVAEEDPG 300
 Qy 301 EPVQAPQPCSSPSNRYHGLSPSSQAPAGAEPTAPLHPRLARHNRGNYRLNMQ 360
 Db 301 EPVQAPQPCSSPSNRYHGLSPSSQAPAGAEPTAPLHPRLARHNRGNYRLNMQ 360
 Qy 361 KHYVGRALRSRLRQAMKQKRRKGGCGGAGVTTKESCFINEOPDHMAAOCPRA 420
 Db 361 KHYVGRALRSRLRQAMKQKRRKGGCGGAGVTTKESCFINEOPDHMAAOCPRA 420
 Qy 421 SEEDTDAVPEPLVPSQPVEVPSIDPTVLPYLSIGPSGQLAETPAEFOALBQLGHOA 480
 Db 421 SEEDTDAVPEPLVPSQPVEVPSIDPTVLPYLSIGPSGQLAETPAEFOALBQLGHOA 480
 Qy 481 FRPQGERAVMRLISGISTLLVPTGAGKSLCYQLPALLYSRSPCLTLVSPLSIMDDQ 540
 Db 481 FRPQGERAVMRLISGISTLLVPTGAGKSLCYQLPALLYSRSPCLTLVSPLSIMDDQ 540
 Qy 541 VSGLPCLKAACIHSGMTRKORSESVLQKIRAAOVNMLTPREALVGAGLPAPAOUPVA 600
 Db 541 VSGLPCLKAACIHSGMTRKORSESVLQKIRAAOVNMLTPREALVGAGLPAPAOUPVA 600
 Qy 601 FACIDEAHCLSQMSHNFRCYLKVCYLREMGVHCFGLTATATRTASDVAAQHLAVAE 660
 Db 601 FACIDEAHCLSQMSHNFRCYLKVCYLREMGVHCFGLTATATRTASDVAAQHLAVAE 660
 Qy 661 EBDLHGPAPVPTNLHLSVSMRDTDOALLTLQGRKPNLDSIIYCNREDETERIALLL 720
 Db 661 EBDLHGPAPVPTNLHLSVSMRDTDOALLTLQGRKPNLDSIIYCNREDETERIALLL 720
 Qy 721 RFLCLAAMVPSGGGAPKTTAAVYHAGMSRERRRQORAFMGOQLRVVYATVAFMGGLR 780
 Db 721 RFLCLAAMVPSGGGAPKTTAAVYHAGMSRERRRQORAFMGOQLRVVYATVAFMGGLR 780
 Qy 781 PDVRAVNLHGLPSPFESYVQAVGRAGRDQPAHCHFLQPOGEDLRELRRHYADSTDEL 840
 Db 781 PDVRAVNLHGLPSPFESYVQAVGRAGRDQPAHCHFLQPOGEDLRELRRHYADSTDEL 840
 Qy 841 AVKRLVQRYFPACTCTTRPSEBQGAVGGERPVKYPQEAOLSHQAAPGPRRYCMGH 900
 Db 841 AVKRLVQRYFPACTCTTRPSEBQGAVGGERPVKYPQEAOLSHQAAPGPRRYCMGH 900
 Qy 901 BRALTIQTLVQALDMPBEAIFETLLCYLELPHHMLELATYTHCCLNPGGPAOLQALA 960
 Db 901 BRALTIQTLVQALDMPBEAIFETLLCYLELPHHMLELATYTHCCLNPGGPAOLQALA 960
 Qy 961 HRCPLAVCLAQOLPEDPGQSSSVFDMVKLVDSMGWELASVRALCOLQMDHEPRTV 1020
 Db 961 HRCPLAVCLAQOLPEDPGQSSSVFDMVKLVDSMGWELASVRALCOLQMDHEPRTV 1020
 Qy 1021 RRGTVLVEFSELAFLHSPGDLTAEBKQICDPLVGRVQAREQALARTFOAFHSV 1080
 Db 1021 RRGTVLVEFSELAFLHSPGDLTAEBKQICDPLVGRVQAREQALARTFOAFHSV 1080
 Qy 1081 AFPSGSPCEQODERSTRKDLGRYREBERGEGMEDAQSPGQARLQMDWQVR 1140
 Db 1081 AFPSGSPCEQODERSTRKDLGRYREBERGEGMEDAQSPGQARLQMDWQVR 1140
 Qy 1141 CDIROPFLSLRPEKSSRAVARIFHGIGSPCYPAQYVQODRRFMRKYLLSLFHALVGLAT 1200
 Db 1141 CDIROPFLSLRPEKSSRAVARIFHGIGSPCYPAQYVQODRRFMRKYLLSLFHALVGLAT 1200
 Qy 1201 BELLOVAR 1208
 Db 1201 BELLOVAR 1208

XX 03-JUN-2004 (first entry)
 DT Human pharmaceutically useful protein SegID 349.
 DE human; cancer; haematopoiesis; thrombosis; anaemia;
 XX cardiovascular disorder; ischaemic heart disease;
 XX acute myocardial infarction; respiratory disease; asthma; pneumonia;
 KW cystic fibrosis; chronic renal failure; glomerulopathy;
 KW gastrointestinal disorder; peptic ulcer; cirrhosis; immune disorder;
 KW HIV infection; systemic lupus erythematosus; endocrine system;
 KW diabetes mellitus; epilepsy; Alzheimer's disease;
 KW amyotrophic lateral sclerosis; skin disorder; psoriasis; bacterial;
 KW fungal; parasitic; viral infection; cytostatic; anticoagulant;
 KW thrombolytic; antihaemic; cardiac; vasotropic; antisthmatic;
 KW antiinflammatory; nephrotropic; antiulcer; hepatotropic;
 KW immunosuppressive; antiallergic; dermatological; antineumatic;
 KW antiaesthetic; antidiabetic; anticonvulsant; neuroprotective; nootropic;
 KW antiparasitic; antibacterial; fungicide; antiparasitic; virucidal;
 KW gene therapy; vaccine.
 XX Homo sapiens.
 OS WO2004020595-A2.
 XX 11-MAR-2004.
 PD 28-AUG-2003; 2003WO-US027107.
 PF 29-AUG-2002; 2002US-0406576P.
 XX 29-AUG-2002; 2002US-0406611P.
 PR 29-AUG-2002; 2002US-0406612P.
 PR 29-AUG-2002; 2002US-0406616P.
 PR 29-AUG-2002; 2002US-0406640P.
 PR 29-AUG-2002; 2002US-0406655P.
 PR 29-AUG-2002; 2002US-0406666P.
 PR 17-SEP-2002; 2002US-0410946P.
 PR 17-SEP-2002; 2002US-0410951P.
 PR 17-SEP-2002; 2002US-0410953P.
 PR 17-SEP-2002; 2002US-0410957P.
 PR 17-SEP-2002; 2002US-0410960P.
 PR 17-SEP-2002; 2002US-0410962P.
 PR 17-SEP-2002; 2002US-0411019P.
 PR 17-SEP-2002; 2002US-0411022P.
 PR 17-SEP-2002; 2002US-0411024P.
 PR 17-SEP-2002; 2002US-0411037P.
 PR 17-SEP-2002; 2002US-0411046P.
 PR 17-SEP-2002; 2002US-0411052P.
 PR 17-SEP-2002; 2002US-0411082P.
 PR 17-SEP-2002; 2002US-0411111P.
 XX (FIVE-) FIVE PRIME THERAPEUTICS INC.
 PA (RIKE-) RIKEN INST PHYSICAL & CHEM RES.
 PA (DNAF-) DNAFORM KK.
 PI Williams LT, Chu K, Lee E, Hestir K;
 XX WPI; 2004-257410/24.
 DR N-PSDB; ADM90747, ADM91165.
 PT New human polynucleotides and polypeptides, useful for diagnosing,
 PT preventing and treating proliferative disorders, immune disorders,
 PT cardiovascular disorders, or bacterial, fungal, parasitic and viral
 PT diseases.
 PS Claim 1, SEQ ID NO 349; 254bp; English.
 XX This invention relates to novel isolated human polynucleotides and the
 CC encoded proteins thereof. Specifically, it refers to proteases, kinases,
 CC phosphatases, secreted and transmembrane proteins, as well as the derived
 CC peptide fragments, which can be used to develop antibodies and screen for
 CC small molecule agonists and antagonists that can modulate their
 CC activities. The present invention describes polypeptides,

CC polynucleotides, vectors and host cells useful for diagnosing, preventing
 CC and treating proliferative disorders, e.g. cancer, disorders of
 CC haematopoiesis such as thrombosis and anaemia, cardiovascular disorders,
 CC e.g. ischaemic heart disease and acute myocardial infarction, respiratory
 CC diseases, e.g. asthma, pneumonia or cystic fibrosis, disorders of the
 CC kidney and urinary tract, e.g. chronic renal failure and glomerulopathy,
 CC gastrointestinal disorders, e.g. peptic ulcer or cirrhosis, immune
 CC disorders, e.g. HIV infection and systemic lupus erythematosus, disorders
 CC of the endocrine system, e.g. diabetes mellitus, central nervous system
 CC disorders, e.g. epilepsy, Alzheimer's disease or anyotropic lateral
 CC sclerosis, skin disorders, e.g. psoriasis, as well as bacterial, fungal,
 CC parasitic and viral diseases. Accordingly, they exhibit many various
 CC activities including cytostatic, anticoagulant, thrombolytic,
 CC antianemic, cardiac, vasotropic, antianasthetic, antiinflammatory,
 CC nephrotropic, antitumor, hepatotropic, immunosuppressive, antiallergic,
 CC dermatological, antineumatic, antiarthritic, antidiabetic,
 CC anticonvulsant, neuroprotective, nootropic, antipsoriatic, antibacterial,
 CC fungicide, antiparasitic and virucidal, such that these polynucleotides
 CC can be used for gene therapy purposes and the development of appropriate
 CC vaccines. This polypeptide is a human protein of the invention.

XX Sequence 1306 AA:

Query Match 94.7%; Score 6085.5; DB 8; Length 1306;
 Best Local Similarity 92.6%; Pred. No. 0;
 Matches 1167; Conservative 2; Mismatches 26; Indels 65; Gaps 6;

QY 1 MERLRDVRERLQWENAFRRQRRPSQDDVENAEPETALREYTTLKRTTGQAGGSR 60
 DB 60 MERLRDVRERLQWENAFRRQRRPSQDDVENAEPETALREYTTLKRTTGQAGGSR 119
 QY 61 SSESLLPAAEEAEPERCMGPHLNRAATKSPPTPGSRQSGVDPYQRLKANLKTLOAG 120
 DB 120 SSESLLPAAEEAEPERCMGPHLNRAATKSPPTPGSRQSGVDPYQRLKANLKTLOAG 179
 QY 121 PALGRRPWLGRASSASTPKPPTGTPVSPFAEKVSDPPQLPEPPRQRLQHLQASIS 180
 DB 180 PALGRRPWLGRASSASTPKPPTGTPVSPFAEKVSDPPQLPEPPRQRLQHLQASIS 239
 QY 181 QRLGSLDPMQLQCHSEVDPFLGAPKACRPDLGSESSQLIPESAVLGPAGSGQPEAS 240
 DB 240 QRLGSLDPMQLQCHSEVDPFLGAPKACRPDLGSESSQLIPESAVLGPAGSGQPEAS 299
 QY 241 AFOEVSIRVSGSPSSGGEKKRWNEEPWESPAAVOGESSQAQPPPEAGAVAVEEDPPG 300
 DB 300 AFOEVSIRVSGSPSSGGEKKRWNEEPWESPAAVOGESSQAQPPPEAGAVAVEEDPPG 359
 QY 301 EPVQAQPPQPCSSPNRYHGLSPSSQAQAKAEGTAPLHIFPRLARHGRNRYRLNMKQ 360
 DB 360 EPVQAQPPQPCSSPNRYHGLSPSSQAQAKAEGTAPLHIFPRLARHGRNRYRLNMKQ 419
 QY 361 KHYVGRALRSRLRKQAMKQKKRKGEGFGGGAIVTTKESFTLNEQDPHMAAQCPRRA 420
 DB 420 KHYVGRALRSRLRKQAMKQKKRKGEGFGGGAIVTTKESFTLNEQDPHMAAQCPRRA 479
 QY 421 SEEDTAVGPEPLVPSPVPEVPSLDPYLPYLSYSGSGQLAETPAEYFOALGOLGHOA 480
 DB 480 SEEDTAVGPEPLVPSPVPEVPSLDPYLPYLSYSGSGQLAETPAEYFOALGOLGHOA 539
 QY 481 FRPGQERAVMRILISGISTLLVPTGAGKSLCYQLPALLYSRSPCLTLVVSPLSLMDQ 540
 DB 540 FRPGQERAVMRILISGISTLLVPTGAGKSLCYQLPALLYSRSPCLTLVVSPLSLMDQ 599
 QY 541 VSGLPCLCAACHSGMTKQKQRESVQKTRAAOVHTLMTLPKLVNAGGILPPAAQLPVA 600
 DB 600 VSGLPCLCAACHSGMTKQKQRESVQKTRAAOVHTLMTLPKLVNAGGILPPAAQLPVA 659
 QY 601 PACIDAHCLSGSHNFRPCYLAECVLEBERMGVHGFGLTAATARTASDVQOHLAAVE 660
 DB 660 PACIDAHCLSGSHNFRPCYLAECVLEBERMGVHGFGLTAATARTASDVQOHLAAVE 719
 QY 661 EPDLHGPAPVPTNLHLVSMMDRDTQALLTLQKGFQNLDSIIYCNREDTERIAALL 720

DB 720 EPDLHGPAPVPTNLHLVSMMDRDTQALLTLQKGFQNLDSIIYCNREDTERIAALL 779
 QY 721 RFLCHAAWPGSGGAPKTTAAVHAGMCSRERRRVSQAFMOGOLR----- 766
 DB 780 RFLCHAAWPGSGGAPKTTAAVHAGMCSRERRRVSQAFMOGOLRGGGCGGLDAGAP 839
 QY 767 -VVAVTVAFGMDLDRPDVAVLHLGLPPSFESYQAVAGRAGRGOPAHCHFLQF----- 820
 DB 840 AHCACAAASGA---FKLRE-LRAGRP-----GRAPLAPRPPTLVLEPPVVAH 886
 QY 821 -----QGEDELRELRHHVHADSTPLAVKLVQVVFPACTCTTRPPSEOGAVGGERP 873
 DB 887 TLMKVALQGEDELRELRHHVHADSTPLAVKLVQVVFPACTCTTRPPSEOGAVGGERP 946
 QY 874 VKYRPOAEQOLSHQAPGPRRVCNGHERALPIQLTVQALDMPDEE----- 918
 DB 947 VKYRPOAEQOLSHQAPGPRRVCNGHERALPIQLTVQALDMPDEE----- 1006
 QY 919 -----ALETLLCYLELHPHNMLELATYTHCRINCBCGPAQLQALAHRCPLAV 968
 DB 1007 PRPLSPALPRALETLLCYLELHPHNMLELATYTHCRINCBCGPAQLQALAHRCPLAV 1066
 QY 969 CLAQQLPEDPGGSGSSVEFDMVKLYDSMGWELASVRALCOQWHEPRTYRGTGVAV 1028
 DB 1067 CLAQQLPEDPGGSGSSVEFDMVKLYDSMGWELASVRALCOQWHEPRTYRGTGVAV 1126
 QY 1029 EFSELAFLHRSRGDLTAEEKDQICDPLGRVQARERQALARLRTTFQAFHSAVPSGCGP 1088
 DB 1127 EFSELAFLHRSRGDLTAEEKDQICDPLGRVQARERQALARLRTTFQAFHSAVPSGCGP 1186
 QY 1089 LEQDEBERSTRKDLIGRYFEEBEGEPGEMDAQPEPQARLDQWEDQVCDIRQFUS 1148
 DB 1187 LEQDEBERSTRKDLIGRYFEEBEGEPGEMDAQPEPQARLDQWEDQVCDIRQFUS 1246
 QY 1149 LRPEKFSRAVARIFHGIGSPCYAQTGQRRTRKTLHSPHALVGLATEBELLQVAR 1208
 DB 1247 LRPEKFSRAVARIFHGIGSPCYAQTGQRRTRKTLHSPHALVGLATEBELLQVAR 1306

RESULT 6
 ADD01202
 ID ADD01202 standard; protein; 1144 AA.
 XX
 AC ADD01202;
 XX
 DT 01-JAN-2004 (first entry)
 XX
 XX
 DE Human nucleic acid-associated protein NAAP-40 SEQ ID NO:40.
 KW human; nucleic acid-associated protein; NAAP; cytosolic;
 KW antiarteriosclerotic; anti-HIV; antiallergic; cerebroprotective;
 KW antiparkinsonian; anticonvulsant; nootropic; neuroprotective;
 KW antinflammatory; ophthalmological; thyromimetic; antiarthritic;
 KW hepatotropic; antibacterial; virucide; protozoacide; antiparasitic;
 KW fungicide; gene therapy; cell proliferative disease; cancer;
 KW atherosclerosis; hepatitis; neurological disorder; Parkinson's disease;
 KW Alzheimer's disease; stroke; epilepsy; developmental disorder;
 KW renal tubular acidosis; anaemia; glaucoma; hypochyroidism;
 KW autoimmune disorder; inflammatory disorder; AIDS; allergy;
 KW atopic dermatitis; arthritis; infection.
 KW
 XX Homo sapiens.
 XX
 OS MO2003054219-A2.
 PN 03-JUL-2003.
 PD 18-DEC-2002; 2002WC-US041115.
 PF 19-DEC-2001; 2001US-0343004P.
 PR 11-JAN-2002; 2002US-0347633P.
 PR 25-JAN-2002; 2002US-0351749P.
 PR 22-FEB-2002; 2002US-0359498P.

XX (INCYTE GENOMICS INC.

PA Azimzai Y, Baughn MR, Becha SD, Borowaky ML, Burford N;
XX Eliott VS, Emerling BM, Forsythe ID, Goryard AE, Griffin JA;
PI Kable R, Khare R, Lai PG, Lee SA, Lee SY, Li JX, Marquis JP;
PI Ramumur J, Richardson TW, Sprague WM, Swarnakar A, Tang YT;
PI Chawla NK, Warren BA, Yue H;

DR WPI, 2003-559157/52.
N-PSDB; ADD01259.

PT New human nucleic acid-associated proteins (NAAP), useful for diagnosing,
PT treating and preventing diseases or conditions associated with the
PT aberrant NAAP expression, e.g. cancer, hepatitis, AIDS, atherosclerosis,
PT infections.

PS Claim 1; SEQ ID NO 40; 405bp; English.

XX The present invention describes human nucleic acid-associated proteins
CC designated NAAP-1 to NAAP-57. The human NAAPs have cytostatic,
CC antiarteriosclerotic, anti-HIV, antiallergic, cerebroprotective,
CC antiParkinsonian, anticonvulsant, nootropic, neuroprotective,
CC antiinflammatory, ophthalmological, rheumatologic, antiarthritic,
CC hepatotropic, antibacterial, virucide, protozoacide, antiparasitic and
CC fungicide activities, and can be used in gene therapy. The NAAP protein
CC and polynucleotide sequences can be used in diagnosing, treating and
CC preventing diseases or conditions associated with the decreased
CC expression or overexpression of NAAP, such as cell proliferative diseases
CC (e.g. cancer, atherosclerosis, hepatitis), neurological disorders
CC (Parkinson's disease, Alzheimer's disease, stroke, epilepsy),
CC developmental disorders (renal tubular acidosis, anaemia, glaucoma,
CC hypothyroidism), autoimmune/inflammatory disorders (AIDS, allergies,
CC atopic dermatitis, arthritis) and infections (e.g. bacterial, viral,
CC parasitic, protozoal, fungal). The present sequence represents human NAAP
CC -40, from the present invention.

XX Sequence 1144 AA;

Query Match 94.2%; Score 6050; DB 7; Length 1144;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1144; Conservative 0; Mismatches 0; Indels 64; Gaps 1;

QY 1 MERLDVBERLQAWERAFRORGRPSODVBAAPETRALYREYTLKRTTGAAGGIR 60
DB 1 MERLDVBERLQAWERAFRORGRPSODVBAAPETRALYREYTLKRTTGAAGGIR 60
QY 61 SSESIPAAAEAPBPRCWPFLNRAATKSPPTPGSRQGSVPDYQRLKANKLKTLOAG 120
DB 61 SSESIPAAAEAPBPRCWPFLNRAATKSPPTPGSRQGSVPDYQRLKANKLKTLOAG 120
QY 121 PALGRRPWFPLGRASSKASTPKPPGTGPVPSFAEKVSDDEPPQLPEPQPRGRLOHLQASIS 180
DB 121 PALGRRPWFPLGRASSKASTPKPPGTGPVPSFAEKVSDDEPPQLPEPQPRGRLOHLQASIS 180
QY 181 QRLGSLDRCMLORCHSEVPDPLGAPACRPDLGSESSQLITGESAVLGPAGSGQPEAS 240
DB 181 QRLGSLDRCMLORCHSEVPDPLGAPACRPDLGSESSQLITGESAVLGPAGSGQPEAS 240
QY 241 AFOEVSIRVGSPOPSSSGEKKRMNEEPWESPAVOQSSQAQPSBAGAVAVEEDPPG 300
DB 241 AFOEVSIRVGSPOPSSSGEKKRMNEEPWESPAVOQSSQAQPSBAGAVAVEEDPPG 300
QY 301 BEVQAQPPQPCSSPNRYHGLSPSSQABAKAEGTAPLHIFPRLARHDGRNVYRLNMQ 360
DB 301 BEVQAQPPQPCSSPNRYHGLSPSSQABAKAEGTAPLHIFPRLARHDGRNVYRLNMQ 360
QY 361 KHYVGRALRSRLAKQAKQKWKKGEGFGGGAATVTTKESCFLEQSDHAAAOCPRA 420
DB 361 KHYVGRALRSRLAKQAKQKWKKGEGFGGGAATVTTKESCFLEQSDHAAAOCPRA 420
QY 421 SEEDTDVAGPEPLVSPQPVPEVPSLDPTVLPLYSLSGSLAETPAEFAQLBQLGHQA 480
DB 421 SEEDTDVAGPEPLVSPQPVPEVPSLDPTVLPLYSLSGSLAETPAEFAQLBQLGHQA 480

DB 421 SEEDTDVAGPEPLVSPQPVPEVPSLDPTVLPLYSLSGSLAETPAEFAQLBQLGHQA 480
QY 481 FRPGGERAVMRILSGISTLLVLPTGAGKSLCTQLPALYSRSPCLTLVSVLSLMDQ 540
DB 481 FRPGGERAVMRILSGISTLLVLPTGAGKSLCTQLPALYSRSPCLTLVSVLSLMDQ 540
QY 541 VSGLPCLKAACIHSGMKRKORESVLQKIRAAQVAVMLTPALVAGAGLPPAOLPVA 600
DB 541 VSGLPCLKAACIHSGMKRKORESVLQKIRAAQVAVMLTPALVAGAGLPPAOLPVA 600
QY 601 FACIDEAHLQSWNSHNPFCYLAVCKVLREBGMVCHFLGTATATARTASDVAOHLAVAE 660
DB 601 FACIDEAHLQSWNSHNPFCYLAVCKVLREBGMVCHFLGTATATARTASDVAOHLAVAE 660
QY 661 EPDLHGPAVPFNHLSVSMRDTQALLTLTGKRFQNLDSIIYCNREDTERIAALL 720
DB 661 EPDLHGPAVPFNHLSVSMRDTQALLTLTGKRFQNLDSIIYCNREDTERIAALL 720
QY 721 RTCLHAAMVPGSGRAPKTTAEYHAGMCSRERRRQAFMGQLRVVAVTAFMGGLDR 780
DB 721 RTCLHAAMVPGSGRAPKTTAEYHAGMCSRERRR----- 755
QY 781 PDVRAVLHGLPPSFESYQAVGRAGDQPAHCHLFLQPGEDURELRHVHADSTDEL 840
DB 756 -----PQGEDURELRHVHADSTDEL 776
QY 841 AVKRLVORVFPACTCTTRPSEOGANGENPVKYPQOEAOLSHQAPPRVCMGH 900
DB 777 AVKRLVORVFPACTCTTRPSEOGANGENPVKYPQOEAOLSHQAPPRVCMGH 836
QY 901 BEALPIQLTVQALDMPBEAIEITLLCYELHPHMEILATTYTHRLNCPGPAQLQALA 960
DB 837 BEALPIQLTVQALDMPBEAIEITLLCYELHPHMEILATTYTHRLNCPGPAQLQALA 896
QY 961 HRCPLAVCLAQQLPEDPQGSSEYFDMVKLVDSMGWELASVRBALCOLQMDHEPRITGV 1020
DB 897 HRCPLAVCLAQQLPEDPQGSSEYFDMVKLVDSMGWELASVRBALCOLQMDHEPRITGV 956
QY 1021 RRGTVLVFESFLAHLASPGDLTAEKQDICTFLYGRVQAEERQALRLRTFOAFHSV 1080
DB 957 RRGTVLVFESFLAHLASPGDLTAEKQDICTFLYGRVQAEERQALRLRTFOAFHSV 1016
QY 1081 APPSGPCLEQDDEBSTRKDLGRYFEESGEPGGMEDAQGPPOGARLODMEDQVR 1140
DB 1017 APPSGPCLEQDDEBSTRKDLGRYFEESGEPGGMEDAQGPPOGARLODMEDQVR 1076
QY 1141 CDIROLSLRPEKSSRAVARIFFIGISPCYPAQVYGQDRRFRKYLHLSFHALVGLAT 1200
DB 1077 CDIROLSLRPEKSSRAVARIFFIGISPCYPAQVYGQDRRFRKYLHLSFHALVGLAT 1136
QY 1201 BEILOVAR 1208
DB 1137 BEILOVAR 1144

RESULT 7
ABM84821
ID ABM84821 standard; protein; 1142 AA.
XX
XX ABM84821;
XX
DE 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic protein SEQ ID NO:5070.
XX
KW gene therapy; human diagnostic and therapeutic polynucleotide; dltip.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX

[illegible]

QY	361	KHYVGRALRSLLAKQAMKQMKRRKKEFGGGGATVTTTKESCFLNEQDHNAAOCBPRA	420
Db	361	KHYVGRALRSLLAKQAMKQMKRRKKEFGGGGATVTTTKESCFLNEQDHNAAOCBPRA	420
QY	421	SEEDTDAVGPBEVLVSPOVPEVPSLDPLVPLVLSIGPSGQLAETPAEYFOALEQGHQ	480
Db	421	SEEDTDAVGPBEVLVSPOVPEVPSLDPLVPLVLSIGPSGQLAETPAEYFOALEQGHQ	480
QY	481	PRPGQERAVMRILSGISTLLVLPTGAGKSLCYQALPALLYSRSPCLTLVSPLLSMDQ	540
Db	481	PRPGQERAVMRILSGISTLLVLPTGAGKSLCYQALPALLYSRSPCLTLVSPLLSMDQ	540
QY	541	VSGLPRLCLKAACISHGKMTKQRESVLQKTRAAQVHYMLTPREALVAGAGLPPAAOLPVA	600
Db	541	VSGLPRLCLKAACISHGKMTKQRESVLQKTRAAQVHYMLTPREALVAGAGLPPAAOLPVA	600
QY	601	PACIDEAHCLSGMSHNPFCYLKVCVLEERAGVHCFGLTATATRTASDVQNHAVAE	660
Db	601	PACIDEAHCLSGMSHNPFCYLKVCVLEERAGVHCFGLTATATRTASDVQNHAVAE	660
QY	661	BEDLHGPAVPVNNMLSVSMRDTQALLTLQKRRFONLDSIIYCNREDTERIAALL	720
Db	661	BEDLHGPAVPVNNMLSVSMRDTQALLTLQKRRFONLDSIIYCNREDTERIAALL	720
QY	721	RTCLHAAMVPGSGGAPKTTAATAYHAGMSERRRRQAPMGQRLVVVATVAFGGLDR	780
Db	721	RTCLHAAMVPGSGGAPKTTAATAYHAGMSERRRR	755
QY	781	PRVRAVHLGLRPPFSYVQANGRAGRDQRAHCHFLDPRQGEDLRLRHVADSTDFL	840
Db	756	-----GDLDELRRHNVADSTDFL	774
QY	841	AVKRLVORVFPACTCTCTPRPSEOEGAVGERPVPKYPROEABOLSHQAPGBRVCMGH	900
Db	775	AVKRLVORVFPACTCTCTPRPSEOEGAVGERPVPKYPROEABOLSHQAPGBRVCMGH	834
QY	901	EEALPQLTLVQALDMPBEALITLLCYLELHPHNMELATTYYHCILNCPGSGAOLQALA	960
Db	835	EEALPQLTLVQALDMPBEALITLLCYLELHPHNMELATTYYHCILNCPGSGAOLQALA	894
QY	961	HNCPLAVLCAOOLPRDPGQSSSVFEDVVKLVDSMGWLASVRLACOLQOMHEPRTGV	1020
Db	895	HNCPLAVLCAOOLPRDPGQSSSVFEDVVKLVDSMGWLASVRLACOLQOMHEPRTGV	954
QY	1021	RKGTGVLVFSELAFLHRSPLDGTABEKQICDFLGRVQAREEQALALRLRTFOAFHSV	1080
Db	955	RKGTGVLVFSELAFLHRSPLDGTABEKQICDFLGRVQAREEQALALRLRTFOAFHSV	1010
QY	1081	AFPSGCPCLCEQDDEERSTRKDLGNYFEEBEGQEPGMEADQEPGQARLQDMEDQVR	1140
Db	1015	AFPSGCPCLCEQDDEERSTRKDLGNYFEEBEGQEPGMEADQEPGQARLQDMEDQVR	1070
QY	1141	CDIRQPLSLRPEEKSSRAVAALFHGIGSPCTPAQVYGGDRRFRWRKYLHLSFHALVGLAT	1200
Db	1075	CDIRQPLSLRPEEKSSRAVAALFHGIGSPCTPAQVYGGDRRFRWRKYLHLSFHALVGLAT	1130
QY	1201	BEILOVAR 1208	
Db	1135	BEILOVAR 1142	
RESULT 8			
AAB20995			
ID	AAB20995 standard; protein; 756 AA.		
XX	AAB20995;		
XX	11-DEC-2000 (first entry)		
XX	Human RecQ4 helicase mutant, mut-2.		
XX	RecQ4 helicase; human; Rothmund-Thomson syndrome; chromosome 8q24.3;		
KW	poikiloderma congenitale; autosomal recessive; skin disorder;		

dermatology; antibody; prenatal diagnosis; gene therapy; mutant; mutein.

XX OS Homo sapiens.

XX PN WO200043522-A1.

XX PD 27-JUL-2000.

XX PF 19-JAN-2000; 2000WO-JP000233.

XX PR 19-JAN-1999; 99JP-00011218.

XX PA (AGEN-) AGENE RES INST CO LTD.

XX PI Kitao S, Shimamoto A, Furuichi Y;

XX DR WPI; 2000-524241/47.

XX DR N-PSDB; AAA72365.

XX PT RecQ4 helicase gene, gene products and antibody, used in the diagnosis

XX and treatment of Rothmund-Thomson syndrome, e.g. by gene therapy.

XX PS Example 3; Page; 115pp; Japanese.

CC The present sequence represents a mutant human RecQ4 helicase, mut-2. The

CC cDNA encoding sequence contains a C to T substitution relative to the

CC cDNA encoding the wild-type RecQ4 helicase (AAA72321), which changes a

CC Gln codon (CAG) to a stop codon (TGA), causing premature truncation of

CC the encoded protein. The invention relates to the genomic DNA sequence of

CC human RecQ4 helicase (AAA72320). Mutations in this gene, located on

CC chromosome 8q24.3, are the cause of Rothmund-Thomson syndrome (also known

CC as poikiloderma congenitale), an autosomal recessive skin disorder

CC principally occurring in females and often accompanied by juvenile

CC cataracts, saddle nose, congenital bone defects, hypogonadism and

CC disturbances in the growth of hair, nails and teeth. The invention also

CC relates to vectors and host cells comprising the human RecQ4 helicase

CC genomic sequence. It additionally encompasses use of the RecQ4 helicase

CC protein as a therapeutic and anti-RecQ4 antibodies as diagnostic agents.

CC The RecQ4 helicase gene and its products, and anti-RecQ4 helicase

CC antibodies are useful in the diagnosis, especially prenatal diagnosis,

CC and treatment of Rothmund-Thomson syndrome. The genomic sequence may

CC especially be used in gene therapy for this condition. Note: The present

CC sequence is not shown in the specification, but is derived from the wild-

CC type human RecQ4 helicase shown on pages 83-92

XX SQ Sequence 756 AA;

Query Match 62.4%; Score 4006; DB 3; Length 756;

Best Local Similarity 100.0%; Pred. No. 2.3e-284;

Matches 756; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MERLTVRERLQAMERAFRRQRGRPPSQDDVEAPETRALYREYTKRTTGAAGGGR 60

DB 1 MERLTVRERLQAMERAFRRQRGRPPSQDDVEAPETRALYREYTKRTTGAAGGGR 60

QY 61 SSESIPAAAEERPEPRCWPMLNRAATKSPQTPGSRGSPVDGQRKATLKTGLQNG 120

DB 61 SSESIPAAAEERPEPRCWPMLNRAATKSPQTPGSRGSPVDGQRKATLKTGLQNG 120

QY 121 PALGRPMPPLGRASSKATPKPKPGTGPVSPFAKTVSDPEQPPEPQPRGRLOHLQASIS 180

DB 121 PALGRPMPPLGRASSKATPKPKPGTGPVSPFAKTVSDPEQPPEPQPRGRLOHLQASIS 180

QY 181 QRLGSLDPGWLORCHSEVDPDLGAPKACRPDLGSEESQLLIFGESAVLGPAGSGCPPEAS 240

DB 181 QRLGSLDPGWLORCHSEVDPDLGAPKACRPDLGSEESQLLIFGESAVLGPAGSGCPPEAS 240

QY 241 AFOEVSIRGSSPOSSSGGEXKRWNEEPESPAYOVQOESSQAGPSEGGAVAVEEDPPG 300

DB 241 AFOEVSIRGSSPOSSSGGEXKRWNEEPESPAYOVQOESSQAGPSEGGAVAVEEDPPG 300

QY 301 BEVQAPQPCSSPSNPRTYGLSPSSQAPAGAEGTAPLHIFPRLARHDRGNVYRLNMKQ 360

DB 301 BEVQAPQPCSSPSNPRTYGLSPSSQAPAGAEGTAPLHIFPRLARHDRGNVYRLNMKQ 360

QY 361 KHYVGRALRSRLRLKQAMKQKRRKKGEFCGGGATVTTKESCPINEOFDHMAAQCPRPA 420

DB 361 KHYVGRALRSRLRLKQAMKQKRRKKGEFCGGGATVTTKESCPINEOFDHMAAQCPRPA 420

QY 421 SEEDTDVAGPEPLVSPPOPVEPVSIDPTVLPLYLSGSPGQLAETPAEYFOALEQSHQA 480

DB 421 SEEDTDVAGPEPLVSPPOPVEPVSIDPTVLPLYLSGSPGQLAETPAEYFOALEQSHQA 480

QY 481 FRPQGERAVMRILISGISLTLVLPFGAGSLCYQLPALVSRSPCLTLVVSPLSLMDQ 540

DB 481 FRPQGERAVMRILISGISLTLVLPFGAGSLCYQLPALVSRSPCLTLVVSPLSLMDQ 540

QY 541 VSGLPPLCAACIHSQMTRKQRESYLQIRAAQVHVMITPEALVAGAGLPAAQLPPVA 600

DB 541 VSGLPPLCAACIHSQMTRKQRESYLQIRAAQVHVMITPEALVAGAGLPAAQLPPVA 600

QY 601 FACIDBAHCLSQWSNRPFCYLRVCKVLREMGVHCFGLTATATRTASDVAOHLAVAE 660

DB 601 FACIDBAHCLSQWSNRPFCYLRVCKVLREMGVHCFGLTATATRTASDVAOHLAVAE 660

QY 661 EPDLHGPAVPPTNLHLVSVMRDITDQALLTLLOGRFQNLDSIIYCNRRREDTERIAALL 720

DB 661 EPDLHGPAVPPTNLHLVSVMRDITDQALLTLLOGRFQNLDSIIYCNRRREDTERIAALL 720

QY 721 RCTLHAANVPSSGSGRAPKTAAHYAAGCSRRRRRV 756

DB 721 RCTLHAANVPSSGSGRAPKTAAHYAAGCSRRRRRV 756

RESULT 9

AAB20994

ID AAB20994 standard; protein; 554 AA.

AC AAB20994;

XX 11-DEC-2000 (first entry)

DE Human RecQ4 helicase mutant, mut-1.

XX RecQ4 helicase; human; Rothmund-Thomson syndrome; chromosome 8q24.3;

KW poikiloderma congenitale; autosomal recessive; skin disorder;

KW dermatology; antibody; prenatal diagnosis; gene therapy; mutant; mutein.

XX OS Homo sapiens.

XX PN WO200043522-A1.

XX PD 27-JUL-2000.

XX PF 19-JAN-2000; 2000WO-JP000233.

XX PR 19-JAN-1999; 99JP-00011218.

XX PI (AGEN-) AGENE RES INST CO LTD.

XX Kitao S, Shimamoto A, Furuichi Y;

XX DR WPI; 2000-524241/47.

XX DR N-PSDB; AAA72365.

XX PT RecQ4 helicase gene, gene products and antibody, used in the diagnosis

XX and treatment of Rothmund-Thomson syndrome, e.g. by gene therapy.

XX PS Example 3; Page; 115pp; Japanese.

CC The present sequence represents a mutant human RecQ4 helicase, mut-1.

CC This cDNA encoding sequence contains a 7 bp deletion relative to the cDNA

CC encoding the wild-type RecQ4 helicase (AAA72321), which alters the

CC reading frame and causes premature truncation of the encoded protein. The

CC invention relates to the genomic DNA sequence of human RecQ4 helicase

CC (AAA72320). Mutations in this gene, located on chromosome 8q24.3, are the

CC cause of Rothmund-Thomson syndrome (also known as poikiloderma
 CC congenitale), an autosomal recessive skin disorder principally occurring
 CC in females and often accompanied by juvenile cataracts, saddle nose,
 CC congenital bone defects, hypogonadism and disturbances in the growth of
 CC hair, nails and teeth. The invention also relates to vectors and host
 CC cells comprising the human RecQ4 helicase genomic sequence. It
 CC additionally encompasses use of the RecQ4 helicase protein as a
 CC therapeutic and anti-RecQ4 antibodies as diagnostic agents. The RecQ4
 CC helicase gene and its products, and anti-RecQ4 helicase antibodies are
 CC useful in the diagnosis, especially prenatal diagnosis, and treatment of
 CC Rothmund-Thomson syndrome. The genomic sequence may especially be used in
 CC gene therapy for this condition. Note: The present sequence is not shown
 CC in the specification, but is derived from the wild-type human RecQ4
 CC helicase shown on pages 83-92

XX Sequence 554 AA;

Query Match 45.5%; Score 2920; DB 3; Length 554;
 Best Local Similarity 100.0%; Pred. No. 5,6e-205;
 Matches 550; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLIDVRERLQAWERAFRRGRSODVAAPEETALREYTTLRITGQAGGR 60
 DB 1 MRLIDVRERLQAWERAFRRGRSODVAAPEETALREYTTLRITGQAGGR 60
 QY 61 SSESILPAAAEAREPERCWPMLNRAATKSPPTPGRSRGSVDYQRLKANKITLQAG 120
 DB 61 SSESILPAAAEAREPERCWPMLNRAATKSPPTPGRSRGSVDYQRLKANKITLQAG 120
 QY 121 PALGRPPWFLGRASSAKATPKPGTGPVPSFAKVSDEPPQLPEOPRRGLQHLQASIS 180
 DB 121 PALGRPPWFLGRASSAKATPKPGTGPVPSFAKVSDEPPQLPEOPRRGLQHLQASIS 180
 QY 181 QRLGSLDPEWMLRCHSEVEDPLGAPKACRPDLSSESSQLITGESVILGPGASQGPBAS 240
 DB 181 QRLGSLDPEWMLRCHSEVEDPLGAPKACRPDLSSESSQLITGESVILGPGASQGPBAS 240
 QY 241 AFOEVSIRVGSFOPSSSGEKKRWNEEPWESPAPVOQESSQAGPPEGAGAVAEEDPPG 300
 DB 241 AFOEVSIRVGSFOPSSSGEKKRWNEEPWESPAPVOQESSQAGPPEGAGAVAEEDPPG 300
 QY 301 EBYQAOPPOPCSPSPNRYHGLSPSSQARAGKAGTAPLHIPPRLANDRGNVRLNMQ 360
 DB 301 EBYQAOPPOPCSPSPNRYHGLSPSSQARAGKAGTAPLHIPPRLANDRGNVRLNMQ 360
 QY 361 KHYVGERALRSRLRKQAKOKKRGKGGCGGATVTTKESCFLEQGDHMAAQCPRA 420
 DB 361 KHYVGERALRSRLRKQAKOKKRGKGGCGGATVTTKESCFLEQGDHMAAQCPRA 420
 QY 421 SEEDTDVAGPEPLVPSQPVEVPSLDPTVLPVLSIGPSGQLAETPAEYFOALEQGHQA 480
 DB 421 SEEDTDVAGPEPLVPSQPVEVPSLDPTVLPVLSIGPSGQLAETPAEYFOALEQGHQA 480
 QY 481 FRPGQERAYMRILISGISTLLVLTPTGAKSLCTQPLALYSRRSPCLTLVSPILSIMDQ 540
 DB 481 FRPGQERAYMRILISGISTLLVLTPTGAKSLCTQPLALYSRRSPCLTLVSPILSIMDQ 540
 QY 541 VSGLPPECLKA 550
 DB 541 VSGLPPECLKA 550

RESULT 10

ABB63317 standard; protein; 1579 AA.

AC ABB63317;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 16743.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PERK) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EM;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL07420.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.

XX Disclosure; SEQ ID NO 16743; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pcc_sequences

XX Sequence 1579 AA;

Query Match 26.8%; Score 1723.5; DB 4; Length 1579;
 Best Local Similarity 28.9%; Pred. No. 1.1e-116;
 Matches 467; Conservative 217; Mismatches 459; Indels 471; Gaps 45;

QY 10 RLQAWERAFRRGRSODVAAPEETALREYTTLRITGQAGGR 51
 DB 10 RLQAWERAFRRGRSODVAAPEETALREYTTLRITGQAGGR 51
 QY 16 RYKWEKOPKKXNGRVSXYDIRDASQELRDSYKMYKLTSPLETLNDVLSBDGYDL 75
 DB 16 RYKWEKOPKKXNGRVSXYDIRDASQELRDSYKMYKLTSPLETLNDVLSBDGYDL 75
 QY 52 -TGQGG-----GIRSS---ESLP----- 66
 DB 52 -TGQGG-----GIRSS---ESLP----- 66
 QY 76 EMSQASDFGVSMIDVSLNBSGPOLPLDISALVGQSSGNLEIPQSVGSEFSNLLDLPN 135
 DB 76 EMSQASDFGVSMIDVSLNBSGPOLPLDISALVGQSSGNLEIPQSVGSEFSNLLDLPN 135
 QY 67 -----AAAEAP-EPRCWGPMLNRAATKSPPTPGRSRGSVDYQRLKANKITLQAG 100
 DB 67 -----AAAEAP-EPRCWGPMLNRAATKSPPTPGRSRGSVDYQRLKANKITLQAG 100
 QY 136 ROVLTNLVNRDENHYIRKEFAVEELPINOAMGLVSKPPRPOPVASKSAFGHGKO- 194
 DB 136 ROVLTNLVNRDENHYIRKEFAVEELPINOAMGLVSKPPRPOPVASKSAFGHGKO- 194
 QY 101 SVPDYQRLKANKITLQAGPALGR--PWLGRASSAKATPKPGTGPVPSFAKVS 157
 DB 101 SVPDYQRLKANKITLQAGPALGR--PWLGRASSAKATPKPGTGPVPSFAKVS 157
 QY 195 --PKAGASLKEPLSAKLFGSSRGFAKRNPKRLKLVSSSSSTGSSVPTDHHLELDF 252
 DB 195 --PKAGASLKEPLSAKLFGSSRGFAKRNPKRLKLVSSSSSTGSSVPTDHHLELDF 252
 QY 158 EPPQLPEOPRRGLQ-----HLQASLQRLSLDPGWLQ-----CHSEV- 199
 DB 158 EPPQLPEOPRRGLQ-----HLQASLQRLSLDPGWLQ-----CHSEV- 199
 QY 253 ETLIRKQOEYKXKQOAIANNPMLASHSKESIKT---LVDDGWLRRNTKENTLDEVEFP 309
 DB 253 ETLIRKQOEYKXKQOAIANNPMLASHSKESIKT---LVDDGWLRRNTKENTLDEVEFP 309
 QY 200 -----DFLAGPK-----ACRPDLSGSESSQLITGESAVL-----GP 230
 DB 200 -----DFLAGPK-----ACRPDLSGSESSQLITGESAVL-----GP 230
 QY 310 AEAANNNGSTKKTNGLANLILSKLPTVKEKVTQAKPDQNALILOELDTMNSMNOXP 369
 DB 310 AEAANNNGSTKKTNGLANLILSKLPTVKEKVTQAKPDQNALILOELDTMNSMNOXP 369
 QY 231 GAGSOGPEASAFQOVSIRVGSFOPSS-----SGGEKRRWNEE----- 267
 DB 231 GAGSOGPEASAFQOVSIRVGSFOPSS-----SGGEKRRWNEE----- 267
 QY 370 DHIINTPPASSQKSAVAPK-NKPPSEQETDSDSVVAASEEQQEYERQLSKKRKIYS 428
 DB 370 DHIINTPPASSQKSAVAPK-NKPPSEQETDSDSVVAASEEQQEYERQLSKKRKIYS 428
 QY 268 -----PWESPAPVOQESS---SQAGP----- 284
 DB 268 -----PWESPAPVOQESS---SQAGP----- 284

```

Db 429 TASGKVEAPAEIENKVEPETETFAQENPDPSADEODATVYENKKDKAKRKOAAK 488
Qy 285 -----PSEGAVAVEEDPGEVQAQPPQ--- 310
Db 489 OKTKPKAEPKTEKAKYKAEKKPKAEKKPRNSKKALAVEBAPPEDEDEQPLMBEDL 548
Qy 311 -----CSSPSNPRYH-----GLSPSQARAGKAGETAP 338
Db 549 KYVLALEAGDITSVRINVOBLEADATQRYITRTFAQPNGLSGSINIRDEKRAAR 608
Qy 339 LHIFRLAHD- RGVYVRLMKQKHYRG- RALRSRLRKQAWKQKR-----KKG 387
Db 609 KLEERIAAGKLNENFTINIOKKKPVKGSVNSFKYKQOBRHKCFVAALSGPMDMG 668
Qy 388 EFGGGGATVTTKESCFLEBQFDHMAAC-----PRP----- 419
Db 669 GCDGG---VL---CFQCGVGHFAQCKVKGDSLPLPSAQLBEDPSPPTLAERQENA 721
Qy 420 -----ASEEDTAVGPEPLVPSQP----- 439
Db 722 SGGAIVVAHSRNTSLPQANAAILOGDELNESEDEQSSGDBEVQGHDPWMSDEMDV 781
Qy 440 -----VPE-----VPSLPT-----VL 451
Db 782 DEBALDAVAEASLSQVSEKASPIKTYVGHKIPPEFLKQAGLDTTASSNRSQHGVK 841
Qy 452 PLYSGSPGQLAETAEVQALFQGHQAFPRQGEAAVNRITISGISTLLVLTGACKSLC 511
Db 842 PLYDLIPDSVODTTPEVLEALHMFQHTFRKQODAIRITSGISLTLTSGSKSLC 901
Qy 512 YQPLALYSRRSPCTLLVSPLLSMDQVSGLPCLKAACJHSGTRQORSVLOKIRA 571
Db 902 YQPLALYSRRKGAITLVISPLVSLMEDQVTPHFLRAKCHTINTQTAQKRIQOMIAN 961
Qy 572 AQVHVLMTPEALV-----GAGGLPPAQLPVAQAFACIDEAHCLSQMSHNRPCYLRYC 625
Db 962 GEIDILVSPBAIVVAGERATFGAL--LQQLPPIAFACIDEAHCLSQMSHNRPCYLRYC 1019
Qy 626 KYLRERMGVHCLGITATRTRTASDVAGHLAAE-EPDLHAPAPPTVTLHLSVMDRPT 684
Db 1020 KYLRNLGVRTVGLTATATLPTLRVSIINHLGISDEREIIIDIPDLVLSVSKDENR 1079
Qy 685 DOALLTLQKRFQUNDIIYCNREDTERIALLRTCLHAAVWSSGGRAPKT---T 740
Db 1080 DAALLQNLNSERFECQSIITCTRDECEIRIAGFRTVQDRREPTQOTKKRKYVMQ 1139
Qy 741 AEAHAGMCSRRERRVORAFMOGOLRVVATVAFQGLDRPVRATLHGLPPSESVYQ 800
Db 1140 AEPYHAGMPASRRRTVQKAFMSNELRIVATIAFGKINKPDIRAVIHYNMFRNESYQ 1199
Qy 801 AVGRAGRGQCPAHCHLFLQPOGEDLRELRRYHADSTDFLAVKRLVQRFPACTCTTRP 860
Db 1200 EIGRAGRGQGLPSHCHFLDAKAGDQSELRHHVYNSIDHVRKLOKIFVCSQSC----- 1254
Qy 861 PSEOGAVGGERPVKPPQGEALQSHQAAPGR-RVCGHGRALPIQITVVALMPREBA 919
Db 1255 --DKASAKRTALPIP-----LEGDPRVHMGCHIGTSVEVTVMELDIPABN 1300
Qy 920 IETLLCYELAHPHMLBELIATYTHCRNLCPGSPALQALAHRCPLAVCLAAQOLPEDEP- 978
Db 1301 ISTLLCYMBLDRWCISVLSAYVMAKVISGCPKYLKHAACECPILMAVALAQIRDKTF 1360
Qy 979 GGGSSVFEEDMKVLVDSMGWELASVRALCOLOMHERTPTGVRGIGVIVSESLAFHNR 1038
Db 1361 KEDSNIIIESVVDIAAGIGMNSGVVYKQDKLEW--VKVNGPKKSPITVSEYDGLFRK 1418
Qy 1039 SPGDLTAEBKQICFLVGRVQAREQALARTFTQAFHSAVAPSCGFCLEQO--DEERS 1097
Db 1419 VPGDTBESEIDNALDTLYTRSKQERTQIQLOLYAHGLAAVAAYSSCGCCNADPQDRG 1478
Qy 1098 TRLKDLGRYFEESGQ-----EPQGMEDAQPEFGQARLODWEQVACDIQRGLSLNR 1151
Db 1479 EQLKAIIVRYFANDYPODLELEIPSNVDP-----ENIIDVHALINMPV 1523

```

```

Qy 1152 EEKSSRAVARIFHGIGSPCAQVYQDNRFRMKYHLHSFALVGLATEELLQ 1205
Db 1524 DNTFTGRNARIIFHGIMSPNPAVIMGR-CKFRAHVKVDFNRILHLAMALIK 1576

RESULT 11
ID AUJ34821
AUJ34821 standard; protein; 610 AA.
XX
AC AUJ34821;
XX
DT 14-FEB-2002 (first entry)
XX
DE E: coli cellular proliferation protein #402.
XX
KW antisense; prokaryotic cellular proliferation protein; antibiotic;
antibacterial; drug design.
XX
OS Escherichia coli.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0251625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyckind JW, Wall D, Trawick JD, Carr GT;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
XX
N-PSDB; AAS52680.
XX
PT New polynucleotides for the identification and development of
antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Example 3; SEQ ID NO 10414; 51bp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence represents an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 610 AA;
XX

```

Query Match 9.8%; Score 631.5; DB 4; Length 610;
 Best Local Similarity 37.6%; Pred. No. 3.1e-37;
 Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

```

QY 474 BOLGQAFRPGQERAWMBILSGISTLVLPTGAGSLCYQLPALYRSRRSPCLTLVVSPL 533
DB 21 EFTFGYQGRPGQEBIITDVLSGRDLVWMTGGSLCYQIALLNG-----LTVVVSPL 76
QY 534 LSLMDQVSGLPP-CLKAACHSGMTRKQRESVLQKIRAAQVVMLTPEALVAGGLPP 592
DB 77 ISLMDQVDOLOANGVAAACLNSTQREQOLEVMTCGRGQIRLILYIAFERIMLDNLFLEH 136
QY 593 AAQLPPVAFACIDEAHCLSSOMSHNRPCLYKVCYLREMGVHNCFLGLTATRTTASV 652
DB 137 LAHNPVLLA-VDEAHCLSSOMSHNRPCLYKVCYLREMGVHNCFLGLTATRTTASV 194
QY 653 AOHLAVERPDLHGPAPVPTNLHLSVMDRDTQALLTLQGRFQNDLSI-----703
DB 195 VRLIG-----LNDPL-----IQIS-SFDRPNIRYLM-----EKFRPDLQMRVYQEQNG 238
QY 704 ---IYCNREDTERIAALLRTCLHAAMVPSGSGRAPKTTAAYAHAGMCSRRRRVQRAF 760
DB 239 KSGIYCNRSRAKVEDTAARL-----OSKGISAAAYHAGLENNVRADVQEKF 284
QY 761 MGOGLRVVAVATVAFPMGGLDRPDVRAVHLGLPSPESYQAVGRAGRDGQPAHCHLFL-- 818
DB 285 QRDDIQLIVAVATVAFPMGGLDRPDVRAVHLGLPSPESYQAVGRAGRDGQPAHCHLFL-- 818
QY 819 -----QPOGEDLRELRH---VHADSTDFLAVKRLV-----QRFPACTC- 855
DB 345 ADMAMLRCLERKPPQGO-LQDIERHKLAMGAFAPAGQCRRLVLANVFBEGQOEPQNGD 403
QY 856 TCTRPPEQEGAVGE 871
DB 404 ICLDPPKQYDGSSTDAQ 419

```

RESULT 12

```

ABU15433
ID ABU15433 standard; protein; 610 AA.
AC ABU15433;
DT 19-JUN-2003 (first entry)
DE Protein encoded by Prokaryotic essential gene #960.
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
OS Escherichia coli.
XX WO200277183-A2.
XX 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-*) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW,
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreysch RA, Xu HH,
XX DR WPI; 2003-029926/02.
XX DR N-PSDB; ACA19303.
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 25; SEQ ID NO 43357; 1766bp; English.
XX

```

CC The invention relates to an isolated nucleic acid comprising any one of
CC the 623 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 610 AA;

```

Query Match          9.8%; Score 631.5; DB 6; Length 610;
Best Local Similarity 37.6%; Pred. No. 3,1e-37;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

QY 474 BOLGQAFRPGQERAWMBILSGISTLVLPTGAGSLCYQLPALYRSRRSPCLTLVVSPL 533
DB 21 EFTFGYQGRPGQEBIITDVLSGRDLVWMTGGSLCYQIALLNG-----LTVVVSPL 76
QY 534 LSLMDQVSGLPP-CLKAACHSGMTRKQRESVLQKIRAAQVVMLTPEALVAGGLPP 592
DB 77 ISLMDQVDOLOANGVAAACLNSTQREQOLEVMTCGRGQIRLILYIAFERIMLDNLFLEH 136
QY 593 AAQLPPVAFACIDEAHCLSSOMSHNRPCLYKVCYLREMGVHNCFLGLTATRTTASV 652
DB 137 LAHNPVLLA-VDEAHCLSSOMSHNRPCLYKVCYLREMGVHNCFLGLTATRTTASV 194
QY 653 AOHLAVERPDLHGPAPVPTNLHLSVMDRDTQALLTLQGRFQNDLSI-----703
DB 195 VRLIG-----LNDPL-----IQIS-SFDRPNIRYLM-----EKFRPDLQMRVYQEQNG 238
QY 704 ---IYCNREDTERIAALLRTCLHAAMVPSGSGRAPKTTAAYAHAGMCSRRRRVQRAF 760
DB 239 KSGIYCNRSRAKVEDTAARL-----OSKGISAAAYHAGLENNVRADVQEKF 284
QY 761 MGOGLRVVAVATVAFPMGGLDRPDVRAVHLGLPSPESYQAVGRAGRDGQPAHCHLFL-- 818
DB 285 QRDDIQLIVAVATVAFPMGGLDRPDVRAVHLGLPSPESYQAVGRAGRDGQPAHCHLFL-- 818
QY 819 -----QPOGEDLRELRH---VHADSTDFLAVKRLV-----QRFPACTC- 855
DB 345 ADMAMLRCLERKPPQGO-LQDIERHKLAMGAFAPAGQCRRLVLANVFBEGQOEPQNGD 403
QY 856 TCTRPPEQEGAVGE 871
DB 404 ICLDPPKQYDGSSTDAQ 419

```

RESULT 13

ABU28286

XX	ABU28286 standard; protein; 609 AA.
XX	ABU28286;
XX	ABU28286;
XX	19-JUN-2003 (first entry)
DE	Protein encoded by Prokaryotic essential gene #13813.
XX	
XX	Antisense; prokaryotic essential gene; cell proliferation; drug design.
OS	Enterobacter cloacae.
PN	MO200277183-A2.
XX	
XX	03-OCT-2002.
XX	
PF	21-MAR-2002; 2002WO-US009107.
XX	
PR	21-MAR-2001; 2001US-00815242.
PR	06-SEP-2001; 2001US-00948992.
PR	25-OCT-2001; 2001US-0342923P.
PR	08-FEB-2002; 2002US-00072851.
XX	06-MAR-2002; 2002US-0362699P.
PA	(ELIT-) ELITRA PHARM INC.
P1	Wang L, Zamudio C, Malone C, Hasselbeck R, Ohlsen KU, Zyekind JW,
P1	Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX	WPI; 2003-029926/02.
DR	N-PSDB; ACP32156.
PT	New antisense nucleic acids, useful for identifying proteins or screening
PT	for homologous nucleic acids required for cellular proliferation to
PT	isolate candidate molecules for rational drug discovery programs.
PS	Claim 25; SEQ ID NO 56210; 1766bp; English.
XX	
XX	The invention relates to an isolated nucleic acid comprising any one of
CC	the 6213 antisense sequences given in the specification where expression
CC	of the nucleic acid inhibits proliferation of a cell. Also included are:
CC	(1) a vector comprising a promoter operably linked to the nucleic acid
CC	encoding a polypeptide whose expression is inhibited by the antisense
CC	nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC	antipeptide or its fragment whose expression is inhibited by the
CC	antisense nucleic acid; (4) an antibody capable of specifically binding
CC	the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC	proliferation or the activity of a gene in an operon required for
CC	proliferation; (7) identifying a compound that influences the activity of
CC	the gene product or that has an activity against a biological pathway
CC	required for proliferation, or that inhibits cellular proliferation; (8)
CC	identifying a gene required for cellular proliferation or the biological
CC	pathway in which a proliferation-required gene or its gene product lies
CC	or a gene on which the test compound that inhibits proliferation of an
CC	organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC	compound's activity; (11) a culture comprising strains in which the gene
CC	product is overexpressed or underexpressed; (12) determining the extent
CC	to which each of the strains is present in a culture or collection of
CC	strains; or (13) identifying the target of a compound that inhibits the
CC	proliferation of an organism. The antisense nucleic acids are useful for
CC	identifying proteins or screening for homologous nucleic acids required
CC	for cellular proliferation to isolate candidate molecules for rational
CC	drug discovery programs, or for screening homologous nucleic acids
CC	required for proliferation in cells other than <i>S. aureus</i> , <i>S. typhimurium</i> ,
CC	<i>K. pneumoniae</i> or <i>P. aeruginosa</i> . The present sequence is encoded by one of
CC	the target prokaryotic essential genes. Note: The sequence data for this
CC	patent did not form part of the printed specification, but was obtained
CC	in electronic format directly from WIPO at
XX	ftp.wipo.int/pub/published_pct_sequences
XX	
XX	Sequence 609 AA;

	Best Local Similarity	35.4%; Pred. No.2.6e-36;			
	Matches	169;	Conservative	73;	Mismatches 152; Indels 84; Gaps 18
QY	474	EQLGHOAFRPOEBAWNRILSGISITLVLPRGAGSLCYQLRALYKRRSPCLTLVNSPL	533		
DB	19	ETFGVOGRPQOEITIEVLEGRDCLVYMPRTGGSGKSLCYQVPAVLVING---	LTVVSPDL	74	
QY	534	LSLMDDOVSG--LPECLKACIHSQWTRKORSVYLQKIAAOVHVLMPREALVGAAGLPR	592		
DB	75	ISLMKDQVDQLLANGVAACLSNSTOTREQOOQVMAQSGCTGQILMYIAPERLMDNFLOH	134		
QY	593	AAOLPRVAFACIDEANCLSQWNSHNERPCYLVKCYLARBMGVNCFGLTATATRRYASDV	652		
DB	135	LAHNHPVLIA--VDEAHCSIQWGHDRPREYALGQ--LRGRPRELPMALTAATDDTTRLDI	192		
QY	653	AQHLVAVEEPLDGHGAPVPTNLHLSV--SMRDTQALITLQKRFQNLDSI-----	703		
DB	193	VRLIG-----LNDP-----YIQVSSFDRNITMYLM-----EKFKPLDQLRLRYOEOBR	235		
QY	704	-----IIVGNREDPERIAALARTCLHAAMVNGSSGRARKTALAYAHAGCSERRRYORA	759		
DB	236	GKSGITVYCNRAKVEDTARLQN-----RGFSAAYAHAGLENIIRADVOEK	281		
QY	760	FMQGLRVVATVAFGMLDRPDYRAVLHLGPRFSFESVVOAGRAGRDGPRAHCHLFLV-	818		
DB	282	FORDDLOLVATVAFGMIKNRPNRVFVNHPRDIPRIEISTYQETGAGRGDGLRAEMLFTPD	341		
QY	819	-----OPQEGDLRELRRH---VHADSTDFLAVKRLV-----QVPRPACTG	855		
DB	342	PADMAWLRRCLEKRGQG--LQDIERHKLNMAGFAFAEQTCRYLVLLNYFGEGRGPRCGNC	400		
QY	856	-TCRRPREEGGAGGERPVK--YRQDAEQLSH-----QAAPRRYCSMGHERLPI	906		
DB	401	DICDPRKQYDGLMDARALSTIYKVNQRFQKGVVEVLRGANNQIRDMGHDK-LPV	457		
RESULT 14					
ABU47859	ID	ABU47859	standard; protein; 609 AA.		
XX	AC	ABU47859;			
XX	XX				
DT	19-JUN-2003	(first entry)			
XX	XX				
DE	Protein encoded by Prokaryotic essential gene #33386.				
XX	XX				
KW	Antisense; prokaryotic essential gene; cell proliferation; drug design.				
OS	<i>Salmomella typhi</i> .				
XX	XX				
PN	WO200277183-A2.				
XX	XX				
PD	03-OCT-2002.				
XX	XX				
PF	21-MAR-2002; 2002WO-US009107.				
XX	XX				
PR	21-MAR-2001; 2001US--00815242.				
XX	XX				
PR	06-SEP-2001; 2001US--00948993.				
XX	XX				
PR	25-OCT-2001; 2001US--0342923P.				
XX	XX				
PR	08-FEB-2002; 2002US--00072851.				
XX	XX				
PR	06-MAR-2002; 2002US--0362699P.				
XX	XX				
PA	(EUIT-) ELITRA PHARM INC.				
XX	XX				
PI	Wang L, Zamudio C, Malone C, Hasegbeck R, Ohlsen KL, ZyeKind JW,				
XX	XX				
PI	Wall D, Trevick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;				
XX	XX				
DR	WPI; 2003-029926/02.				
XX	XX				
DR	N-PSDB; ACA51729.				
XX	XX				
PT	New antisense nucleic acids, useful for identifying proteins or screening				
XX	XX				
PT	for homologous nucleic acids required for cellular proliferation to				
XX	XX				
PT	isolate candidate molecules for rational drug discovery programs.				
XX	XX				


```
Db      81 ISLMDQVLDQLANGVAAACLNSTQSREQLVMAACRTGQIRLXYIAERLMDNFDLH 140
OY      593 AAQLPPVAFACIDEAHCLUSQWSHNPFCYLKVCYLREBMGVHCEGLTATATRTASDV 652
Db      141 LAHNPVLLA-VDEAHCTISQWGHDFRPEYALGQ-LRQRFALPFMALATATADFTTRDI 198
OY      653 AQHLVAEERPDHGPAPVPTNLHLSVSMRDTDQALLTLQGRFQNLDSI----- 703
Db      199 IRLLG-----LNDPL-----IQIS-SFDRPNIRYMLM---BKRPDLQLMRYVQEORG 242
OY      704 ---IYCNRREBDTERIAALLRTCLHAAMVPGSGRAPKTTABAYHAGMCSRRRRVQRAF 760
Db      243 KSGIITYCNSRAVEDTAAALQ-----SRG-----ISAAYHAGLENAIRADVQEKF 288
OY      761 MOGOLRVVAVATVAFGWLDRPDVRAVLHGLPPSFESYVQAVGRAGRDGOPAHCHLFIOP 820
Db      289 QRDLOIVVATVAFGWIKNKPNVRFVHFDIPNIESYYQETGRAGRDGLPAEAMLFYDP 348
OY      821 -----QGEDLRELRRHVTADSTDFLAVKRLV-----QRVPACTC-T 856
Db      349 ADMAMLRRCLEKPPAGQLQDIERHKLNAMGAFEAQTCRRLVLNYPGEGROEPCGNCDI 408
OY      857 CTRPPSEOEG 866
Db      409 CLDPPKQYDG 418
```

Search completed: March 19, 2005, 01:04:38
Job time : 190 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2005, 01:01:38 ; Search time 165 Seconds
(without alignments)
2419.962 Million cell updates/sec

Title: US-09-889-325-4

Perfect score: 6424

Sequence: 1 MERLADVNERIQAWERARRR.....HLSFALVGLATRELLQVAR 1208

Scoring table:

Gapop 10.0, Gapext 0.5

Searched: 1401741 seqs, 330541175 residues

Total number of hits satisfying chosen parameters: 1401741

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6424	100.0	1208	11 US-09-889-325-4	Sequence 4, Appli
2	765	11.9	874	16 US-10-427-963-121617	Sequence 121617,
3	632.5	9.8	607	15 US-10-374-077-75	Sequence 75, Appl
4	631.5	9.8	610	9 US-09-815-242-10414	Sequence 10414, A
5	631.5	9.8	610	15 US-10-282-122A-43357	Sequence 43357, A
6	621	9.7	370	9 US-09-753-143-77	Sequence 77, Appl
7	619	9.6	609	15 US-10-282-122A-56210	Sequence 56210, A
8	609.5	9.5	609	15 US-10-282-122A-75783	Sequence 75783, A
9	609.5	9.5	615	9 US-09-815-242-11747	Sequence 11747, A
10	607.5	9.5	615	15 US-10-282-122A-74973	Sequence 74973, A
11	606	9.4	608	15 US-10-282-122A-59929	Sequence 59929, A
12	593	9.2	610	15 US-10-282-122A-78025	Sequence 78025, A
13	589.5	9.2	600	15 US-10-282-122A-69214	Sequence 69214, A

14	581.5	9.1	620	15 US-10-282-122A-76978	Sequence 76978, A
15	578.5	9.0	632	15 US-10-282-122A-67293	Sequence 67293, A
16	578	9.0	619	9 US-09-815-242-11085	Sequence 11085, A
17	578	9.0	619	15 US-10-282-122A-58271	Sequence 58271, A
18	554.5	8.6	589	15 US-10-282-122A-42447	Sequence 42447, A
19	552	8.6	590	15 US-10-282-122A-60691	Sequence 60691, A
20	549	8.5	597	15 US-10-282-122A-72864	Sequence 72864, A
21	547.5	8.5	589	9 US-09-815-242-10803	Sequence 10803, A
22	545	8.5	580	9 US-09-815-242-4959	Sequence 4959, A
23	545	8.5	712	15 US-10-282-122A-11915	Sequence 11915, A
24	545	8.5	712	15 US-10-282-122A-66476	Sequence 66476, A
25	544.5	8.5	390	15 US-10-282-122A-61441	Sequence 61441, A
26	544	8.5	341	15 US-10-369-493-8990	Sequence 8990, Ap
27	542.5	8.4	1955	14 US-10-293-604-3	Sequence 3, Appli
28	541	8.4	588	16 US-10-437-963-14145	Sequence 141345, A
29	528	8.2	590	15 US-10-282-122A-57618	Sequence 57618, A
30	526	8.2	592	15 US-10-282-122A-70676	Sequence 70676, A
31	525.5	8.2	1328	15 US-10-374-077-76	Sequence 76, Appl
32	524.5	8.2	658	14 US-10-156-761-12385	Sequence 12385, A
33	518.5	8.1	715	15 US-10-282-122A-67821	Sequence 67821, A
34	517	8.0	653	15 US-10-425-114-59459	Sequence 59459, A
35	515	8.0	827	15 US-10-282-122A-53151	Sequence 53151, A
36	514	8.0	593	15 US-10-282-122A-70264	Sequence 70264, A
37	513	8.0	593	9 US-09-815-242-12306	Sequence 12306, A
38	511	8.0	601	15 US-10-282-122A-48328	Sequence 48328, A
39	510	7.9	592	15 US-10-282-122A-71773	Sequence 71773, A
40	508	7.9	335	15 US-10-369-493-10345	Sequence 10345, A
41	507.5	7.9	720	15 US-10-282-122A-48551	Sequence 48551, A
42	506.5	7.9	410	15 US-10-443-108-2	Sequence 2, Appli
43	505.5	7.9	478	15 US-10-369-493-4682	Sequence 2682, Ap
44	505	7.9	473	15 US-10-425-114-42472	Sequence 42472, A
45	504	7.8	648	15 US-10-282-122A-51285	Sequence 51285, A

ALIGNMENTS

RESULT 1					
US-09-889-325-4					
Sequence 4, Application US/09889325					
Publication No. US20040224312A1					
GENERAL INFORMATION:					
APPLICANT: AGENE Research Institute, Co., Ltd.					
TITLE OF INVENTION: Gene causative of Rothmund-Thomson syndrome					
FILE REFERENCE: AI-003PCT					
CURRENT APPLICATION NUMBER: US/09/889,325					
CURRENT FILING DATE: 2001-07-13					
PRIOR APPLICATION NUMBER: JP 1999-11218					
PRIOR FILING DATE: 1999-01-19					
NUMBER OF SEQ ID NOS: 46					
SOFTWARE: PatentIn Ver. 2.0					
SEQ ID NO 4					
LENGTH: 1208					
TYPE: PRT					
ORGANISM: Homo sapiens					
US-09-889-325-4					
Query Match					
Best Local Similarity 100.0%; Score 6424; DB 11; Length 1208;					
Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MERLADVNERIQAWERARRRQRRRPSDDVEAAPEETRALYREYRTIKRTTGAAGGLR	60		
DB	1	MERLADVNERIQAWERARRRQRRRPSDDVEAAPEETRALYREYRTIKRTTGAAGGLR	60		
QY	61	SSSELPAAEBAPEPCGPHLNRAATKSPQTPGRSHQSGVPPYGGRLKNTLKTLOAG	120		
DB	61	SSSELPAAEBAPEPCGPHLNRAATKSPQTPGRSHQSGVPPYGGRLKNTLKTLOAG	120		
QY	121	PALGRPWPPLGRASSKASTPKPGTGPVPSFAEKVSDPPQLPBPQPRGLQHLQASLS	180		
DB	121	PALGRPWPPLGRASSKASTPKPGTGPVPSFAEKVSDPPQLPBPQPRGLQHLQASLS	180		

181 QRLGSLDPGMLORCHSEVDFLGA PKACRPDLGSEESQLLIPGESAVLPGAGSQGPBAS 240
181 QRLGSLDPGMLORCHSEVDFLGA PKACRPDLGSEESQLLIPGESAVLPGAGSQGPBAS 240
QY 241 AFOEVSIRVGSQOPSSSGGKKRMNEEMESPAPVOQOESSQAGPPEGA GA VAEEDPPG 300
Db 241 AFOEVSIRVGSQOPSSSGGKKRMNEEMESPAPVOQOESSQAGPPEGA GA VAEEDPPG 300
QY 301 BEVOAOPPCSSPSNPRYHGLSPSSQAGAKAEGTAPLHPRLARHNGVYVRLNMQO 360
Db 301 BEVOAOPPCSSPSNPRYHGLSPSSQAGAKAEGTAPLHPRLARHNGVYVRLNMQO 360
QY 361 KHYVGRALRSRLRKQAWKOKMKKKGECFGGGGATVTTKESCFLENEFDHMAAQCPRPA 420
Db 361 KHYVGRALRSRLRKQAWKOKMKKKGECFGGGGATVTTKESCFLENEFDHMAAQCPRPA 420
QY 421 SEEDTDVAGPEPLVPSPOVPEVPSLDPTVPLYSLSGPGQLAEPTAEVFOALBOLGHQA 480
Db 421 SEEDTDVAGPEPLVPSPOVPEVPSLDPTVPLYSLSGPGQLAEPTAEVFOALBOLGHQA 480
QY 481 PRPGOERAVMRLISGISTLLVLTPTGAGKSLCYQLPALYSRBSPCLTIVVSPILSLMDQO 540
Db 481 PRPGOERAVMRLISGISTLLVLTPTGAGKSLCYQLPALYSRBSPCLTIVVSPILSLMDQO 540
QY 541 VSGLPPECLKAACISHSMTKRORESYLQKIRAAQVHVLMTPREALVGAGGLPRAAQLPVA 600
Db 541 VSGLPPECLKAACISHSMTKRORESYLQKIRAAQVHVLMTPREALVGAGGLPRAAQLPVA 600
QY 601 FACIDBAHCLSQMSHNFRCYLCVLCYLBKRGVHFCGLTATATRTTSDVAOHLAVAE 660
Db 601 FACIDBAHCLSQMSHNFRCYLCVLCYLBKRGVHFCGLTATATRTTSDVAOHLAVAE 660
QY 661 BEDLHGPAVPTNHLVSMMDRDTDOALLTLLOGKSFQMLDSIIICNREDTERIAALL 720
Db 661 BEDLHGPAVPTNHLVSMMDRDTDOALLTLLOGKSFQMLDSIIICNREDTERIAALL 720
QY 721 RFLCNAHVPGSGGGRAPKTTAAEYHAGMCSRERRVQORAFMOGQLHVVAATVAFGGLDR 780
Db 721 RFLCNAHVPGSGGGRAPKTTAAEYHAGMCSRERRVQORAFMOGQLHVVAATVAFGGLDR 780
QY 781 PRVRAVTLHGLPPSFSSYVOAVGRARDQOPACHFLPOGEBDLRELRHVADSTDL 840
Db 781 PRVRAVTLHGLPPSFSSYVOAVGRARDQOPACHFLPOGEBDLRELRHVADSTDL 840
QY 841 AVKRLVQRYFPACTCTCTRPSPBOGAVGGERVPRKYPROBAQLSHQAAPGRVCMGH 900
Db 841 AVKRLVQRYFPACTCTCTRPSPBOGAVGGERVPRKYPROBAQLSHQAAPGRVCMGH 900
QY 901 ERALPIQLTVQALDMEBAIETLLCYLBLEHPHMELELATYTHCRINCSPGPAQLQALA 960
Db 901 ERALPIQLTVQALDMEBAIETLLCYLBLEHPHMELELATYTHCRINCSPGPAQLQALA 960
QY 961 HRCPLAVCLAOQLPEDDPGQSSSVFEDVVKVDSNGWELASVRAALCOLQMDHNERTV 1020
Db 961 HRCPLAVCLAOQLPEDDPGQSSSVFEDVVKVDSNGWELASVRAALCOLQMDHNERTV 1020
QY 1021 RRGTVLVEFSEIAFLRSPGDLTAEEKQICDPLGRVQAREQALRLRFTFOAFHSY 1080
Db 1021 RRGTVLVEFSEIAFLRSPGDLTAEEKQICDPLGRVQAREQALRLRFTFOAFHSY 1080
QY 1081 AEPSCGPELEQODEESTRLKDLIGRYFEEBEGQEGMEDAOGPEPGAQLDQMEDQYR 1140
Db 1081 AEPSCGPELEQODEESTRLKDLIGRYFEEBEGQEGMEDAOGPEPGAQLDQMEDQYR 1140
QY 1141 COIROGLSLRPEKSSRAVARIFHGIGSPCAVYQODRRFRWKYHLHSFHALVGLAT 1200
Db 1141 COIROGLSLRPEKSSRAVARIFHGIGSPCAVYQODRRFRWKYHLHSFHALVGLAT 1200
QY 1201 BEILOVAR 1208
Db 1201 BEILOVAR 1208

RESULT 2
US-10-437-963-121617
; Sequence 121617, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Mu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Bardazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 121617
; LENGTH: 874
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(874)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_24624C.1.pep
US-10-437-963-121617

Query Match 11.9%; Score 765; DB 16; Length 874;
Best Local Similarity 24.1%; Pred. No. 3.2e-41;
Matches 274; Conservative 162; Mismatches 380; Indels 320; Gaps 39;
QY 110 KANLKGTLQAGPALGRPPWRPLGRASSKASTPKP---GTGVPSTAEKYSDERPOLPEQO 166
Db 13 EGSLLSDVASAPRRRRSP---PRPAPRRPPRRPPPKHRRPAPRTKPKTKPRPAASABA 69
QY 167 PRPGRLQHLQALSGRLSLDP---GWLORCHSEVDFLGA PKACRPDLGSEES---QLLTP 222
Db 70 PAPPPTTLRAAALS---DPHGLAAR-----IAAGSLTAASGASSSFRRLVQ 116
QY 223 GESAVLGPAGSQGEASAFQEVSIKVGSPQSSSGGKKRMNEEMESPAPVOQOESSQA 282
Db 117 SNBPSPDRTATATAASSAPSEVPASAARPPRTAA-----TDAPQTRPRVHP 165
QY 283 GPPSAGAVAEEDPPGEVQAPPPQPCSSPSNPRYHGLSPSSQAGAKAEGTAPLHIF 342
Db 166 NSVSEVMAASAABOP-----KDAKGG----- 187
QY 343 PRLAHDRNGVYVRLNMQOKHYVGRALRSRLRKQAWKOKMKKKGECFGGGATVTTKES 402
Db 188 -----SEGNFRLNIN---NGYGRKRTFKSQAKSTKCSWKRQ---RAAGT----- 229
QY 403 CFLENEFDHMAAQCPRPASEEDTDVAGPEPLVPSPOVPEVPSLDPTVPLYSLSGSGQL 462
Db 230 -----FRSQGDESGDLVAEALLREKQASD-----SVLEAVE-----SV 264
QY 463 AETPAEVFOALBOL-----GHOAFRPGOERAVMRLISGISTLLVLTPTGAGKSLCYQLPAL 517
Db 265 REDPSE---ONLKSLLNAAVGHDSFPOGQLEAIQOIVAGESTMLVLTPTGAGKSLCYQVPM 322
QY 518 LYSRBSPCLTIVVSPILSLMDQOVGGLPPECLKAACISHSMTKRORESYLQKIRAAQVHVL 577
Db 323 IL-----PGLTLVSPILSLMDQVQLKRLPAFLPGGLASSQTSDEPHDTQLRPAAGTAYL 378
QY 578 MLTPREALVGAGGLPRAAQLPVAFACTIDBAHCLSQMSHNFRCYLCV---CKVLRERMGVAC 636
Db 379 FVSPERFLNEERLULFROTLPISTIVAIIDBAHCLSQMSHNFRCYLCVLRSLRKLKLVQC 438
QY 637 FLGLTATATRTTASDVAOHLAVAEEDDLHGPAVPTNHLVSMMDRDTDOALLTLLOGKR 696

```

Db      439 ILAMRATATQTLBEELMABLEIPSD-NLQTOIRNLOLSTSTSNRLKDLMLLKSP 497
Qy      697 FONDLSIIYCNRRREDTERIALALRTCLHAAMVPSGGAPKTTAAVYAGCRRRRV 756
Db      498 FVDMRSIIYVCKFOAETDFVSKYL-CDN-----NITAKSYSHGLIKRSHV 543
Qy      757 QAPFQOGLRVVAVYAFQMGIDRPVRAVHLGLPSPFESYQAVGRAGROGPAHCHL 816
Db      544 QELFCSNKRIF-----SY-----NF 559
Qy      817 FLQPGCEDLRELRHVADSTDFLAVKRLVORVPACTCTCTRPSEOGAVGERPVK 876
Db      560 TEQF-----FLGSDGVGYAMSKLYQIF-----SSENTTG----- 590
Qy      877 YPQAEQSLQAPQPRVRCVGHERRALPIQLTVQALDPEEAIEITLLCYLELHPHMLE 936
Db      591 -----CIG-----SLAKEITSRKFIDEVLTITLQLEIGDQOYIR 627
Qy      937 LLATYTHCRNCPGPAQOLALHRCPLAVC-----LAQQLPDEPGOSSSVPEPMVL 992
Db      628 LIPQSVTCTL-----YFHKTSPLLADKIDILRSVLANSEMKDGHVFDIPRI 676
Qy      993 VDSMGELASVRRALCOLQMDHEPRTGVRGTGLVEFS-ELAFHLRSPG----- 1041
Db      677 ANDLKITMNEV-----FDHLHK-----LKFSGEISFLKDPAYCYVILMRD 718
Qy      1042 DLTAEEKQICDPLVGRVQARERQALRLRTQAFHSAVAPSC-----GPCLE 1090
Db      719 DENALSAN-----LTKWLSEVESKISKLDAMF-ALANPAVAGCKRTGSCSGQHPICIQ 772
Qy      1091 OODERSIRLKLGLRYFEFEEOGEBGCMEDAGPEPGARIQDMEDVRCIRQPLSLR 1150
Db      773 KKLIME-----YFSKDDGTSEND-----RTQLOKSSPFLQADIKVFIQSN 812
Qy      1151 PEEKSSRAVARIFGIGSPCYPAQVQODRRFRMRYLHLSFHALVGLATEELLQV 1206
Db      813 SRAKTPRAVARIMGSISSPAFSPVTWGN-HFKGRVYEDVPLVMEAKALIVLU 867

```

```

RESULT 3
US-10-374-077-75
; Sequence 75, Application US/10374077
; Publication No. US20040006779A1
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; Yu, Chang-En
; Oshima, Junko
; Mulligan, John T.
; Schellenberg, Gerald D.
; TITLE OF INVENTION: ANTIBODIES AGAINST GENE PRODUCTS RELATED TO
; WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Intellectual Property Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/374,077
; FILING DATE: 25-Feb-2003
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 100107, 401D1

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 607 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 75:
US-10-374-077-75

Query Match      9.8%; Score 632.5; DB 15; Length 607;
Best Local Similarity 37.6%; Pred No. 9,2e-33;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

Qy      474 EQLGQAPRPGGERAWMILSGISTLLVLPFGAGSLCYQVALLYSRRSPCLTVVSP 533
Db      18 EFGYQFRPQGEIIDTVLSGRDCLVMPFGSKSLCYQIPALLNG---LTVVSP 73
Qy      534 LSLMDQVSGLP-CLKAACIHSGMTKQRESVYLQKIRAAQVHVMLTPALVAGAGLP 592
Db      74 ISLMQDVQLQANGVAAACINSTYTRBQOLEVMTGCRGTQIRLLYIAPERLMDNFLH 133
Qy      593 AALPVPVAFACIDEAHCLSGNSHNPFCYLRYCKVLRRMGVHCFGLTATRTASDV 652
Db      134 LAHNPVLLA-VDEHNCISQMGHDPREYALGQ-LRQRFPLPMALATADDTTRQDI 191
Qy      653 AQHLVAEBPDLHGPAVPYTMHLSSVMDRDTQALLTLQKRFQNDLSI----- 703
Db      192 VRLIG-----LNDPL-----IQIS-SFDRPIRYMLM-----EKFKPLDQLMRVVOEORG 235
Qy      704 ---IYCNRRREDTERIALALRTCLHAAMVPSGGAPKTTAAVYAGCRRRRVQAF 760
Db      236 KSGIYCNRRREDTERIALALRTCLHAAMVPSGGAPKTTAAVYAGCRRRRVQAF 281
Qy      761 MGGQLRVVAVYAFQMGIDRPVRAVHLGLPSPFESYQAVGRAGROGPAHCHLFL-- 818
Db      282 QRDQLQIVAVYAFQMGINKRVRVVFHFDIPRNIESTYQEGRAGROGDLPRMALFDP 341
Qy      819 -----QPGEDLRELRH---VHADSTDFLAVKRLV-----QVFPACTC- 855
Db      342 ADMAWLRCLBEKPGQ-LQDIERHKLNAMGAFABAQTCRRRLVLTNYRGEGRQEGCNGD 400
Qy      856 TCTRPSEOGAVGGE 871
Db      401 ICLDPKQYDGSSTDAQ 416

```

```

RESULT 4
US-09-815-242-10414
; Sequence 10414, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; Applicant: Chlesen, Kari L.
; Applicant: Zyskind, Judith W.
; Applicant: Wall, Daniel
; Applicant: Trawick, John D.
; Applicant: Carr, Grant J.
; Applicant: Yamamoto, Robert T.
; Applicant: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26

```

```
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 10414
/ LENGTH: 610
/ TYPE: PRF
/ ORGANISM: Escherichia coli
US-09-815-242-10414

Query Match      9.8%; Score 631.5; DB 9; Length 610;
Best Local Similarity 37.6%; Pred. No. 1.1e-32;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

QY 474 BOLGHAAPRQGERAVMRLISGISTLVLPFGAGKSLCYQLPALLYSRSPCLTIVVSP 533
DB 21 EFTFGYQRRPQGEELIDYVLSGRDCLVMPFGGKSLCYQIPALLNG---LTVVVSPL 76
QY 534 ISLMDQVSGLP-CLKACIHSGMTKQRESVLOKIRAAQVHVMTLTPREALVAGAGLPP 592
DB 77 ISLMDQVQDLOANGVAAACLNSTOTREQOLEVMTCGRGQIRLYIAPERLMDNPLEH 136
QY 593 AAQLPPVAFACIDEAHCLSQSHNRPCLRYCKYLREMGVHCFGLTATATRTTASV 652
DB 137 LAHWPVLLA-VDEAHCSQWGHDPREYALGQ-LRQFPPLPFMALATATADTTRODI 194
QY 653 AOHLAVAEPPDLHGAPVPTNLHLSVSMRDTQALLTLLOGKRQNDLSI----- 703
DB 195 VLLG-----LNDPL-----IQIS-SFDRPNIRYMLM-----EKFPDLQLMRYVQEQRG 238
QY 704 ---IYCNREDTERIALLRCTCLHAAMVPGSGGAPKTTAAVYAGMCSRERRRVORAF 760
DB 239 KSGIITCNRAKVEDTARL-----QSKGISAAAYAGLENNVRADVQEKF 284
QY 761 MGGQLRVVATVAFMGGLDRPDVRAVLHLGLPPSFESYQVAGRAAGDQPAHCHLFL-- 818
DB 285 QRDLDIVATVAFMGINKPNVRPVVHFDIPRNIESYQETGRAGRDGLPEAMLFYDP 344
QY 819 -----QPGEDLRELRH---VHADSTDFLAVKRLV-----QVFPACGC- 855
DB 345 ADMAMLRCLERKPPQO-LQDIERHKLAMGAFAEAQTCRRLVLTNYFEGEQEPCGNC 403
QY 856 TCTRPSEQEGAVGGE 871
DB 404 ICLDPPKQYDGSSTDAQ 419

RESULT 5
US-10-282-122A-43357
/ Sequence 43357, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Foreytch, R.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
```

```
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO: 43357
/ LENGTH: 610
/ TYPE: PRF
/ ORGANISM: Escherichia coli
US-10-282-122A-43357

Query Match      9.8%; Score 631.5; DB 15; Length 610;
Best Local Similarity 37.6%; Pred. No. 1.1e-32;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

QY 474 BOLGHAAPRQGERAVMRLISGISTLVLPFGAGKSLCYQLPALLYSRSPCLTIVVSP 533
DB 21 EFTFGYQRRPQGEELIDYVLSGRDCLVMPFGGKSLCYQIPALLNG---LTVVVSPL 76
QY 534 ISLMDQVSGLP-CLKACIHSGMTKQRESVLOKIRAAQVHVMTLTPREALVAGAGLPP 592
DB 77 ISLMDQVQDLOANGVAAACLNSTOTREQOLEVMTCGRGQIRLYIAPERLMDNPLEH 136
QY 771 ISLMDQVQDLOANGVAAACLNSTOTREQOLEVMTCGRGQIRLYIAPERLMDNPLEH 136
QY 593 AAQLPPVAFACIDEAHCLSQSHNRPCLRYCKYLREMGVHCFGLTATATRTTASV 652
DB 137 LAHWPVLLA-VDEAHCSQWGHDPREYALGQ-LRQFPPLPFMALATATADTTRODI 194
QY 653 AOHLAVAEPPDLHGAPVPTNLHLSVSMRDTQALLTLLOGKRQNDLSI----- 703
DB 195 VLLG-----LNDPL-----IQIS-SFDRPNIRYMLM-----EKFPDLQLMRYVQEQRG 238
QY 704 ---IYCNREDTERIALLRCTCLHAAMVPGSGGAPKTTAAVYAGMCSRERRRVORAF 760
DB 239 KSGIITCNRAKVEDTARL-----QSKGISAAAYAGLENNVRADVQEKF 284
QY 761 MGGQLRVVATVAFMGGLDRPDVRAVLHLGLPPSFESYQVAGRAAGDQPAHCHLFL-- 818
DB 285 QRDLDIVATVAFMGINKPNVRPVVHFDIPRNIESYQETGRAGRDGLPEAMLFYDP 344
QY 819 -----QPGEDLRELRH---VHADSTDFLAVKRLV-----QVFPACGC- 855
DB 345 ADMAMLRCLERKPPQO-LQDIERHKLAMGAFAEAQTCRRLVLTNYFEGEQEPCGNC 403
QY 856 TCTRPSEQEGAVGGE 871
DB 404 ICLDPPKQYDGSSTDAQ 419

RESULT 6
US-09-753-143-77
/ Sequence 77, Application US/09753143
/ Patent No. US20020102550A1
/ GENERAL INFORMATION:
/ APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
```

GRONEN
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
OF BLOOM'S SYNDROME
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/753,143
FILING DATE: 02-Jan-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/175,828
FILING DATE: 1998-10-20
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/65
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 370
TYPE: AMINO ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: <Unknown>
DESCRIPTION: OTHER NUCLEIC ACID
HYPOTHETICAL: YES
ANTI-SENSE: NO
FEATURES:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-09-753-143-77
Query Match 9.7%; Score 621; DB 9; Length 370;
Best Local Similarity 40.0%; Pred. No. 2.7e-32;
Matches 154; Conservative 57; Mismatches 110; Indels 64; Gaps 12;

474 BOLGAQRRPQGERAVMRILSGISTLLVLPFGAGSLCYQLPALYSRSPCLTLVVSPL 533
4 EFTFGQRRPQGERIITLVLSGRDCLVMPFGGSLCYQVPAVLNG---LTVVVSPL 59
534 LSLMDQVSGLEPP-CLKAACTHSGMTRKQRESVLOKIRAQVHVMLTPEALVAGAGLPP 592
60 ISLMDQVDOQLANVAAACLNSTOTREQOLEVMTGCRGTQIRLVIAPERLMDNFDLH 119
593 AAQLPPVAFACIDEAHCLSQSHNFRPCYLRVCYLREMGVHCFGLTATATRTTASDV 652
120 LAHNPVLLA-VDEAHCLSQSHNFRPCYLRVCYLREMGVHCFGLTATATRTTASDV 177
653 AQLHVAEPEDLGAPVPTNLHLVSMDRDTDAALLTLQSKRPQNDLSI-----703
178 VRLTG-----LNDPL-----IQIS-SFDRPNIRVTLM-----EKFPDQLARVYQEQNG 221
704 ---IYCKRRETERIAALRTCLAAVVPVSGGAPKTTATAYAGMSRRRRVQAF 760
222 KSGIITVCSRAKVEDTAAAL-----OSKGISAAVYHAGLENNYRAVQOEKF 267
761 MGGQLRVVVATVAFGAGLDRPDVRAVHLGLPFPFESVYQAVGARGDQPAHCHLFL-- 818

DB 268 QRDDIQTIVATVAFGAGLDRPDVRAVHLGLPFPFESVYQAVGARGDQPAHCHLFL 327
QY 819 -----QPGEDIRLELRH 831
DB 328 ADMWLRRLCEKPEQOQ-LQDIERH 351
RESULT 7
US-10-282-122A-56210
; Sequence 56210, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Hasebeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56210
; LENGTH: 609
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-10-282-122A-56210
Query Match 9.6%; Score 619; DB 15; Length 609;
Best Local Similarity 35.4%; Pred. No. 7e-32;
Matches 169; Conservative 73; Mismatches 152; Indels 84; Gaps 18;

474 BOLGAQRRPQGERAVMRILSGISTLLVLPFGAGSLCYQLPALYSRSPCLTLVVSPL 533
19 EFTFGQRRPQGERIITLVLSGRDCLVMPFGGSLCYQVPAVLNG---LTVVVSPL 74
DB 19 EFTFGQRRPQGERIITLVLSGRDCLVMPFGGSLCYQVPAVLNG---LTVVVSPL 74
QY 534 LSLMDQVSG-LEPCLKAACTHSGMTRKQRESVLOKIRAQVHVMLTPEALVAGAGLPP 592
75 ISLMDQVDOQLANVAAACLNSTOTREQOLEVMTGCRGTQIRLVIAPERLMDNFDLH 134
DB 75 ISLMDQVDOQLANVAAACLNSTOTREQOLEVMTGCRGTQIRLVIAPERLMDNFDLH 134
QY 593 AAQLPPVAFACIDEAHCLSQSHNFRPCYLRVCYLREMGVHCFGLTATATRTTASDV 652
135 LAHNPVLLA-VDEAHCLSQSHNFRPCYLRVCYLREMGVHCFGLTATATRTTASDV 192

```

QY 653 AOHIAVAEERDLHGAPVPNTNHLST-SWDRPTDOLITLLGCRKONDSI----- 703
Db 193 VRLIG-----LNDP-----YIOVSSDREPIRKIML-----EKKRPDOLIRYOEOR 235
QY 704 ----IYCNRRDTERIALITLTCLHAAMVPSGGRAPPTTAAYAHAGCSRRERRVORA 759
Db 236 GKSGLIYCNRSRAKVEDTAARLON-----RGFSAAAYAGLENNHRAVDQEK 281
QY 760 FMQGLRVVAIVAAEGMGLDRPDVAVLHLGLPPSPESVYVANGRAGPSQAPAHCHFL- 818
Db 282 FORDDLQIVATVAAGMGINKPNVRKVFHFDIPIRNTESYQOETGRAGDGLPREAMLFTYD 341
QY 819 -----QOGBEDLREILRRH---VHADSTDFLAKVRLV-----QEVFPACTC 855
Db 342 PADMAMLRRCLEKEPQOG-LDIDIERHKLNAMGAFAPQTCRRRLVLNYPGEGRQEPQNC 400
QY 856 -TCTRPSPBQEGAVGGERFVRK-YRQGEABQLSH-----QAAGPRRVCMGHERALPI 906
Db 401 DICLDPKQYDGLMDARKALSTTYRVNQPFGMGVVEVLGANNQIRIDMGHKH-LPV 457

```

RESULT 8
US-10-282-122A-75783
; Sequence 75783, Application US/10282122A
; Publication No. US20040029129A1

```

1  APPLICANT: Wang, Liansun
2  APPLICANT: Zamudio, Carlos
3  APPLICANT: Malone, Cheryl
4  APPLICANT: Haselbeck, Robert
5  APPLICANT: Ohlsen, Kari
6  APPLICANT: Zykkind, Judith
7  APPLICANT: Wall, Daniel
8  APPLICANT: Trawick, John
9  APPLICANT: Carr, Grant
10 APPLICANT: Yamamoto, Robert
11 APPLICANT: Forsyth, R.
12 APPLICANT: Xu, H.
13 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
14 FILE REFERENCE: EPI7PA.034A
15 CURRENT APPLICATION NUMBER: US/10/282,122A
16 CURRENT FILING DATE: 2003-02-20
17 PRIOR APPLICATION NUMBER: 60/131,078
18 PRIOR FILING DATE: 2000-03-21
19 PRIOR APPLICATION NUMBER: 60/206,848
20 PRIOR FILING DATE: 2000-05-23
21 PRIOR APPLICATION NUMBER: 60/207,727
22 PRIOR FILING DATE: 2000-05-26
23 PRIOR APPLICATION NUMBER: 60/230,335
24 PRIOR FILING DATE: 2000-09-06
25 PRIOR APPLICATION NUMBER: 60/230,347
26 PRIOR FILING DATE: 2000-09-09
27 PRIOR APPLICATION NUMBER: 60/242,578
28 PRIOR FILING DATE: 2000-10-23
29 PRIOR APPLICATION NUMBER: 60/253,625
30 PRIOR FILING DATE: 2000-11-17
31 PRIOR APPLICATION NUMBER: 60/257,931
32 PRIOR FILING DATE: 2000-12-22
33 PRIOR APPLICATION NUMBER: 60/267,636
34 PRIOR FILING DATE: 2001-02-09
35 PRIOR APPLICATION NUMBER: 60/269,308
36 PRIOR FILING DATE: 2001-02-16
37 Remaining Prior Application data removed - See File Wrapper or PALM.
38 NUMBER OF SEQ ID NOS: 78614
39 SOFTWARE: PatentIn version 3.1
40 SEQ ID NO 75783
41 LENGTH: 609
42 TYPE: PRT
43 ORGANISM: Salmonella typhi
44 US-10-282-122A-75783

```

Best Local Similarity 37.7%; Pred. No. 2.9e-31;
Matches 162; Conservative 58; Mismatches 137; Indels 73; Gaps 14;

QY	474	EQLGHOARPPQOEAVNRMLSGISFTLVLPFGAGSKSLCYOYBALYXSRSPCGLTVVSL	533
Db	19	ETFGIQOEPRQOEALIDTALSGRDLVMPFTGGSLCTQIPALLDGG---LTVVSVSL	74
QY	534	LSLMDDDVSG-LPRCLAKACIHSGMTRKORESVLOKIRAAOVHVMILPREALVGAGLPR	592
Db	75	ISLMKDQYDQOLANGVAALCINSTQSRQOELVMAAGCRNGIRLLIYIAFERMLDNLFDH	134
QY	593	AAQLEPRVAFACIDENHCISOWSHNRPCYLKVCXKLEEMSGNHCGLGTATATRPASDV	652
Db	135	LAHNPVLLA-VDEAHCIISQWCHDRPREVAALQG-LRORFPLPFMALATATADDTTRODI	192
QY	653	AQHLVAEERPLHGAPVPYMLMLSVSMDRDTQALTLLOQKRFONLDSI-----	703
Db	193	IRLLG-----LNDPL-----IQIS-STDRPIRITML-----EKKEPDLQMLRVYQEORG	236
QY	704	---IINCRRDTERIALLTCTLHAAMPVSGSGAPKPTTAAYAHGCMCSRRRRYQRAF	760
Db	237	KSGLIITCSNRAKVEEDTARLQ-----SRG-----ISAAAYAHGLENATRADQEKF	282
QY	761	MQGQRRVVAVATYAFEMGLDRPVRVAVLHGLRPPESVYQAVAGRGDQPAHCHFLDP	820
Db	283	QRDDQLVAVATYAFEMGINKPRVNRVVFHEDIPRNIESTYQETGGRGRDGLPHEAMLFTYP	342
QY	821	-----QGEDLRELRHHADSTDFLAVKRLV-----QRVFPACTC-T	856
Db	343	ADMAWLRRCLEBKPAQQLODIERHKLNMGMGAPAAQOTGRVLVILNYFGEGROEPCGNCDI	402
QY	857	CTRPSPSEOG 866	
Db	403	CLDPEKQYDG 412	

RESULT 9
US-09-815-242-13747

; Sequence 13747, Application US/09815242
; Patent No. US20020061569A1

```

/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari L.
/ APPLICANT: Zyskind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of Essential Genes in
/ TITLE OF INVENTION: Prokaryotes
/ FILE REFERENCE: ELITRA.011A
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ CURRENT FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 13747
/ LENGTH: 615
/ TYPE: PRT

```

ORGANISM: *Salmonella typhi*
US-09-815-242-13747

Query Match 9.5%; Score 609.5; DB 9; Length 615;
Best Local Similarity 37.7%; Pred. No. 3e-31;
Matches 162; Conservative 58; Mismatches 137; Indels 73; Gaps 14;

QY 474 EOLGHOAPRPGOERAVMRILSGISTLLVPTGAGSKLCYQLPALYSRRSPCLTLVSP 533
DB 25 EFGGYQOFRPGOBAIIDTLASGRDCLVWPTGGKSLCYQIPALLDG---LTVVSP 80
QY 534 ISLMDQVSG-LPPLKAACTHSGMTRKORRESYLQKIRAOVHVM/LTPEALVGAGLPP 592
DB 81 ISLMDQVQDLANGVAACLNSTOSREGQLEVMAGCRGQIRLYIAERLMDLNF 140
QY 593 AAOLPPVAFACIDEAHCSOMSHNRPCTLYRVCYLREMGVHCFGLTATATRTASDV 652
DB 141 LAHNPVLLA-VDEAHCTISQMGHDFRPEYALGQ-LRQRPALPFWALATATADTTRODI 198
QY 653 AOHVAEERPDHGPAPVPTNLHLSVMDRTDQALLTLQKRPONLDSI----- 703
DB 199 IRLG-----LNDPL-----IQIS-SFDRPNIRYMLM---SKKPLDQLMKRYVOEQRG 242
QY 704 ---IYCNREDTERIALLRCTLHAAMVPGSGGAPKTTAAVHAGMCSRERRVQRA 760
DB 243 KSGIYCNSRAKVEDTARLQ-----SRG-----ISAAYHAGLENALRADVQEK 288
QY 761 MOGQLRVVATVAFMGDLRDPVRAVLHGLPPSFESYQAVGRAGDQPAHCHFL 820
DB 289 QRDLOIVATVAFMGINKPVRFVHFDIPRNIESYQOEGRAGRDLPAEAMLFYDP 348
QY 821 -----QGEDLRELRHVHADSTDFLAVRLV-----QVPPACTC-T 856
DB 349 ADMAMLRCLBEKPAQQLDIERHKLNAMGAFAEACTRRLVLTNVPFGRQEPGNC 408
QY 857 CTRPSEBOG 866
DB 409 CLDPPKQYDG 418

RESULT 10
US-10-282-122A-74973
; Sequence 74973, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patencin version 3.1
; SEQ ID NO 74973
; LENGTH: 615
; TYPE: PRF
; ORGANISM: *Salmonella typhimurium*
US-10-282-122A-74973

Query Match 9.5%; Score 607.5; DB 15; Length 615;
Best Local Similarity 37.7%; Pred. No. 4e-31;
Matches 162; Conservative 58; Mismatches 137; Indels 73; Gaps 14;

QY 474 EOLGHOAPRPGOERAVMRILSGISTLLVPTGAGSKLCYQLPALYSRRSPCLTLVSP 533
DB 25 EFGGYQOFRPGOBAIIDTLASGRDCLVWPTGGKSLCYQIPALLDG---LTVVSP 80
QY 534 ISLMDQVSG-LPPLKAACTHSGMTRKORRESYLQKIRAOVHVM/LTPEALVGAGLPP 592
DB 81 ISLMDQVQDLANGVAACLNSTOSREGQLEVMAGCRGQIRLYIAERLMDLNF 140
QY 593 AAOLPPVAFACIDEAHCSOMSHNRPCTLYRVCYLREMGVHCFGLTATATRTASDV 652
DB 141 LAHNPVLLA-VDEAHCTISQMGHDFRPEYALGQ-LRQRPALPFWALATATADTTRODI 198
QY 653 AOHVAEERPDHGPAPVPTNLHLSVMDRTDQALLTLQKRPONLDSI----- 703
DB 199 IRLG-----LNDPL-----IQIS-SFDRPNIRYMLM---SKKPLDQLMKRYVOEQRG 242
QY 704 ---IYCNREDTERIALLRCTLHAAMVPGSGGAPKTTAAVHAGMCSRERRVQRA 760
DB 243 KSGIYCNSRAKVEDTARLQ-----SRG-----ISAAYHAGLENALRADVQEK 288
QY 761 MOGQLRVVATVAFMGDLRDPVRAVLHGLPPSFESYQAVGRAGDQPAHCHFL 820
DB 289 QRDLOIVATVAFMGINKPVRFVHFDIPRNIESYQOEGRAGRDLPAEAMLFYDP 348
QY 821 -----QGEDLRELRHVHADSTDFLAVRLV-----QVPPACTC-T 856
DB 349 ADMAMLRCLBEKPAQQLDIERHKLNAMGAFAEACTRRLVLTNVPFGRQEPGNC 408
QY 857 CTRPSEBOG 866
DB 409 CLDPPKQYDG 418

RESULT 11
US-10-282-122A-59929
; Sequence 59929, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A

Query Match	Similarity	9.4%	Score 606;	DB 15	Length 608;
Best Local	Similarity	35.5%	Pred. No. 5e-31;		
Matches	Conservative	70;	Mismatches 157;	Indels	80; Gaps 18;
QY	474	EOLGQAEPQGEAAVWRILSGISITLVLVPGAGKSLCYOLPALLYSRRSPCLTLVVSPL			533
Db	19	ETFGYQGRPPGEETLLETALTEGRDCLVMPPTGGKSLCYOVALVMGQ---LTVVASPL			74
QY	534	LSIMDDQVSG-LRPCLKAACIHSGMTKRORESVLQKIRPAOVHVMLTPEALVGAGLPP			592
Db	75	ISIMDDQVDOLLANVAAACINSTOSRQDQGEVMAAGCSGGVRLLYIAPERLMDNFLSH			134
QY	593	AAQLPVAACIDEAHCLISQWSHNPFCYLAVCKYLREMGVHCFGLTATATRTTASDV			652
Db	135	LANW-NLMLAVDEAHCTISQWGHPRPYAALGQ-LRQRMQIPFMAALTATADDTTRDI			192
QY	653	AOHLVAREPDLHGPAVPYTMILHSVSDMDRDTQALLTLQKRCFONDSI-----			703
Db	193	VRLLG-----LNDPL-----IQVS-SEDRPIRITMLM-----EKKFPDQLMKRYQDDRG			236
QY	704	---IYCNRREDTERIALLTCLHAAMPVSGSGAPKTTAEAVYHAGCSRRRRVORAF			760
Db	237	KSGLIYCSNRSKVEEDTAAQLQ-----SRG-----ISAAAYAGLENDVRAVQSEKF			282
QY	761	MGGQLRVVVAATYAFMGAGDRPDYRAVLHGLSPSESVYQVANGRAGDGPANHGLFLQD			820
Db	283	QDDQLQIVAVYAFMGAGINKPVRKVFVHEPDI PRNIESYIYQETGRAGDGLPRAEMLFYDP			342
QY	821	-----QGEDIRELRRH---VHAQSTDEFLAVKRLV-----QRFVPACTG-T			856
Db	343	ADMAWLRRCLEEKPAGPLQDIERHKILNMGAFAEAQOTCRRLVLINVFGEGRQEPGNCIDI			402
QY	857	CTRPSEDEGANGGRRVPVK-YPPDEABQLSH-----QAAPRRRVCMGHEALPLI			906
Db	403	CLDPKQQIDGLMDARKALSTTYRVNQRFGMGVVEVLKAGANNQRIREMGKH-LFV			457

RESULT 12

US-10-282-122A-78025

; Sequence 78025, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

```

; LENGTH: 610
; TYPE: PRT
; ORGANISM: Yersinia pestis
US-10-282-122A-78025

```

Best Local Similarity 33.3%; Pred. No. 3.5e-30;
Matches 169; Conservative 71; Mismatches 172; Indels 96; Gaps 18

474 EGTCHQAEEDPCQEPNMPITSGTSTLTITDTGAGKSLCYOLPAITVSRPSPCTITVVSPI 533

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

334 LSLIMDDQVSGLEFC-LNHACIHSGMINKKESVDKNIKQNVAVDITLLENVCHOCUL 332

Db 75 ISLMKDQVDTLAYGVGAGCLNSSQTRBQLAVMDGCRSGQIKLLYIAPEKLVMESEFLDQ 134

593 AAQLPVAACIDEAHCLSQWSHNERPCYLKRVCKLRERMGVHCFGLTATATRTASDV 652

Db 135 LYQWRP-ALLAVDEAHCSQWGHDFRPEY-RALGQLKQRFDDLPIVIALTATADEATRCDI 192

653 AOHLAVAEEDPLHGAPVPTNLHLSVSMDDPTDOALLTLLOGKRFQNLDSI----- 703

```

      : | : | : | : | :
103 MBT - - - TOT$-SEDPBNTPYTYV---EKFKPI.DOI.WREVONOBG 236

```

760 E 9TH AVE

[illegible]

DB 237 KSG11CNSKAKVEDI LARLU-----SRG-----ESVAKH1HAGDDNENUCQ*YENR 202

761 MGGRLRVVATVAFGMGLDRPDVRAVLHLGLPPSFESYVQAVGRAGRDGQPAHCHLELP 820

Db 283 QRRDLQVWATVAFGMGINKPNVRFVHEDIPTIESYYQETGRAGRDGLPAEAVLLYDP 342

QY 821 QG-----EDLRLRRHVADSTDFLAVKRYVORF-----PACTC-T 856
DB 343 ADMAMLRRLERKPRAGAOQDIENKUNMGAFABEATCTCRVLVLTNGEGKQOPCGNCBI 402
QY 857 CTRPSEOGGAVGERPVP-KYRPOEAOISH-----QAAPRRVCMGHERALPIQITV 910
DB 403 CLDPFRKYDGLADAGALSCYVRVGFGLYVEVLRGANNORIRIEMGHDK-----LSV 457
QY 911 QALDMPREBAITLCTLEHPIHPIWLELL 938
DB 458 YGIGR-----EQTHEWVSVL 473

RESULT 13
US-10-282-122A-69214
Sequence 69214, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zysek, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA 034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR FILING DATE: 2000-03-21
PRIOR FILING DATE: 2000-03-21
PRIOR FILING DATE: 2000-05-23
PRIOR FILING DATE: 2000-05-23
PRIOR FILING DATE: 2000-05-26
PRIOR FILING DATE: 2000-05-26
PRIOR FILING DATE: 2000-09-06
PRIOR FILING DATE: 2000-09-06
PRIOR FILING DATE: 2000-09-09
PRIOR FILING DATE: 2000-10-23
PRIOR FILING DATE: 2000-11-27
PRIOR FILING DATE: 2000-11-27
PRIOR FILING DATE: 2000-12-22
PRIOR FILING DATE: 2000-12-22
PRIOR FILING DATE: 2001-02-09
PRIOR FILING DATE: 2001-02-09
PRIOR FILING DATE: 2001-02-16
PRIOR FILING DATE: 2001-02-16
Remaining prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patent version 3.1
SEQ ID NO: 69214
LENGTH: 600
TYPE: PRT
ORGANISM: Proteus mirabilis
US-10-282-122A-69214

Query Match 9.2%; Score 589.5; DB 15; Length 600;
Best Local Similarity 32.2%; Pred. No. 5, 8e-30;
Matches 184; Conservative 69; Mismatches 151; Indels 167; Gaps 24;

QY 474 BOLGQAPRPGOBRVMTLSGISTLVYPTGAGSLCYOLALLYSRSPCLTIVSVSL 533
DB 10 EFTFGYQGFPGGQBITDITITRDCLVNPTGGGSLCYQIPALLDGG---LTVVSVSL 65
QY 534 ISLMDQVSGLRPCL---KACISGMRKQRESVYLQKIRAAQVHVMLTPEALVAGGL 590

DB 66 ISLMKQVQDL---CLHGIDAAFLNSTQSRDEQVQMGORGBIKLYIAPERLMMESFL 123
QY 591 PPAQQLPVAFACIDBAHCLISQSNRPFCYLAVCYLBERGVNHCFLGLTATATRTS 650
DB 124 HHLVQWQP-ALLAVDEAHCTISQMGHDFRPEY-RGICLRLQVLPDVPIALTAADNTTTR 181
QY 651 DVAGHLVABEEDPLHGPAVPNTLHVSMDTDQDALTLLOQKRFQNDST----- 703
DB 182 DIINQLALR-----TFL---VHIS-SFDENRIRYTV---EKKPLDQIMLFIHQ 225
QY 704 -----IYCNREDETERIALLRCTLHAAMVDSGGGRAPKTAAYHAGVCSRRRRVOR 758
DB 226 KKGSGIYCNRSRKYETAEFL-----GKRG-----LSIAVHAGMDIATRAKVD 271
QY 759 AFMOGQLRVVATVAVGKMLDPRDRAVTLHLGLPFSFSTVQAVGAGDGPACHFL 818
DB 272 AFORDLOIVAVTVAFGKINRPNVRFVHFPDIPRNIESTYQETGAGHDGPAQAVLFY 331
QY 819 OPGEDRLRLRRHVADSTDFLAVKRYVORFPACTCTRPSEOGGAVGERPVPKYP 878
DB 332 DP--ADMAMLR-----CLDEKPESEOKAI----- 354
QY 879 POEAOISHQAAAPRRVCMGHERALPIQITVQALDMPREBAITLCTLEHPIHPIWLELL 938
DB 355 --ESHKUNMGAFABAQTC---RRL-----VLLNYF----- 380
QY 939 ATTYTHCRNLCPGPAQOLALHRCPLAVCLAQQLPEPQGGSSVBERDMVKLDSMG 998
DB 381 -----GHRQQA---CGNCDICL-----DPRK-----QYD---GLVDA--- 407
QY 999 ELASVRALCOLQWHEHPTGVRGTVLVE 1029
DB 408 -----OKALSCYI---RTGQRFGLGYVE 428

RESULT 14
US-10-282-122A-76978
Sequence 76978, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zysek, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA 034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR FILING DATE: 2000-03-21
PRIOR FILING DATE: 2000-03-21
PRIOR FILING DATE: 2000-05-23
PRIOR FILING DATE: 2000-05-23
PRIOR FILING DATE: 2000-05-26
PRIOR FILING DATE: 2000-05-26
PRIOR FILING DATE: 2000-09-06
PRIOR FILING DATE: 2000-09-06
PRIOR FILING DATE: 2000-09-09
PRIOR FILING DATE: 2000-10-23
PRIOR FILING DATE: 2000-11-27
PRIOR FILING DATE: 2000-11-27
PRIOR FILING DATE: 2000-12-22
PRIOR FILING DATE: 2000-12-22
PRIOR FILING DATE: 2001-02-09
PRIOR FILING DATE: 2001-02-09
PRIOR FILING DATE: 2001-02-16
PRIOR FILING DATE: 2001-02-16
Remaining prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patent version 3.1
SEQ ID NO: 76978
LENGTH: 600
TYPE: PRT
ORGANISM: Proteus mirabilis
US-10-282-122A-76978

Query Match 9.2%; Score 589.5; DB 15; Length 600;
Best Local Similarity 32.2%; Pred. No. 5, 8e-30;
Matches 184; Conservative 69; Mismatches 151; Indels 167; Gaps 24;

QY 474 BOLGQAPRPGOBRVMTLSGISTLVYPTGAGSLCYOLALLYSRSPCLTIVSVSL 533
DB 10 EFTFGYQGFPGGQBITDITITRDCLVNPTGGGSLCYQIPALLDGG---LTVVSVSL 65
QY 534 ISLMDQVSGLRPCL---KACISGMRKQRESVYLQKIRAAQVHVMLTPEALVAGGL 590

```

? PRIOR FILING DATE: 2001-02-09
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? Remaining Prior Application data removed - See File Wrapper or PALM.
? NUMBER OF SEQ ID NOS: 78614
? SOFTWARE: Patentin version 3.1
? SEQ ID NO 76978
? LENGTH: 620
? TYPE: PRT
? ORGANISM: Vibrio cholerae
? OS-10-282-122A-76978

```

Query Match	9.1%;	Score	581.5;	DB	15;	Length	620;
Best Local Similarity	33.0%;	Pred. No.	2e-29;				
Matches	160;	Conservative	74;	Mismatches	171;	Indels	80;
						Gaps	14

QY	456	GGPSQGLAETPAVAVQALQOLGHOAPRPOEAVNMRILGSLITVLPTAGKSLQOLP	51.5
Db	15	MAESSALPTERRVLH--EVFGIQOFRVGOEVEITALMRDSVLNMFPGKSLQOLP	72
QY	516	ALLVRSRSPCLTLVVSPLSLMDDOVSGDPR-CLKAACIHSGMTKRORESVLQIKIRAOV	574
Db	73	ALVEEG---VTLVISPLISIMKDQDOQLKANGVAEVSUNTLAREELIAYNRMHAGOL	128
QY	575	HYLMITPRELVAGAGSLPRAAOPLRVVAFACIDENHCISQWNSHNRPCITRYCKXILREMGV	634
Db	129	KLLVSPBEVLTAEPETERLSH--PLMIMAVDENHCISQWHDHRRPREYASIGQ--LKORFEN	186
QY	635	HCFLGTLTATRTTRTASDVAAQHLAVAEERPLNIGRAPPTVNLHLSVSMRDITDOALLTLQOG	694
Db	187	VPVMALTATADADATRHMDIQOLQNLNPHOYLGSFDRPNIRKYMVEKKHQPVSQ--VIRYLET	245
QY	655	KRFQULDSIIITYCNREDTERIALAALRTCLHAAWVSGSGRAKTALEAHNMGCSERR	75
Db	246	ONGQ---CGIITCGSKKKVEMLTEKJ--C-----GNHNR-----AASHAGNDADBERA	288
QY	755	RQOARFMQOGLRVVVATVAFGMGLDRPDVAVALHLGSLPSESUYVAVAGVAGRDGCZPANC	814
Db	289	WVQEAFFORDDLQIVATVAFGMGINKNPNRVFVHFDIPRIEISUYOETGAGNRDGLPABA	348
QY	815	HLFLPOQEGDLELRHNHVAADSTDFLAVKRLVORVPRPACTCTSTRPESBEGAVGSGRPV	874
Db	349	MWLYDP--ADMWMLRREML-----DEKPD 365	
QY	875	PKCPROEABOLSHOAPGPRRVLC-----MGERALPIQLTVQALDMP-----EEAI	920
Db	370	GNQKQVESHSKLTAMGFAFAAQTCRQVULNVPBYVDKPCGNCDICLDPRKRPDTEBAR	422
QY	921	ETLLC 925	
Db	430	KALSC 434	

```

RESULT 15 1
US-10-282-122A-67293
: Sequence 67293, Application US/10282122A
: Publication No. US20040029123A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Liangsu
: APPLICANT: Zamudio, Carlos
: APPLICANT: Malone, Cheryl
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari
: APPLICANT: Zyskind, Judith
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John
: APPLICANT: Carr, Grant
: APPLICANT: Yamamoto, Robert
: APPLICANT: Forsyth, R.
: APPLICANT: Xu, H.
: TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
: FILE REFERENCE: ELITRA.034A
: CURRENT APPLICATION NUMBER: US/10/282.122A

```

```

CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 67293
LENGTH: 632
TYPE: PRT
ORGANISM: Pasteurella multocida
US-10-282-122A-67293

```

Query Match	9.0%;	Score 578.5;	DB 15;	Length 632;
Best Local Similarity	32.8%;	Pred. No. 3.3e-29;		
Matches 149; Conservative	76;	Mismatches 142;	Indels 87;	Gaps 15;

```

QY 462 LAEPFAEYFOALBOUQHAFRQGBERAWMRLISGISTVLVLPFGAGKSLCYOUPALYLX 521
QY 22 LKQTLADVLHNV -FGYOSFRKGQEEVLDATLMDKDSLVATMAGKSLCYOUPALCFEG 79
Db 522 RSPCLTVVSPILSLMDQVSGS-LRPSCLKAACIHSGMTREKORESVLOKIRAAOVHMLT 580
QY 80 -LTVISPLISLMDKQVOQDLANGIEADYLSNGTFTGBOQOVNKLMSGTLKLLYVS 135
Db 561 PEALVAGAGLPPAOLPVYAFACIDBANCLSGWSMNFRCYLRVCKYLBERNCVHF--- 637
QY 136 PEKWVTSFPFLHISIC-KVSPVALIDBANCIQWGHDFREYQLOGL-----KSCFPHA 188
Db 638 --IGLTATRTRTASDVNOHLVAEERDLGCAVPYMLHLSV-SMDRDTQDLATLLOG 694
Db 189 PIMALTATRADTATRODILRHL-----NLOSP-----HYVIGSDRNRINRTLV---- 231
QY 695 KRFOULDSI-----IYCNREDTERIALLTCLHAAVPGSGRAPRTTAE 742
Db 232 EKFKEMEOJCFPVLOQKKSGIYCNRSKVRIRIESLNN-----KGVSAQ 277
QY 743 AYHAEKRERRRYORAFMOGOLRVVTVAFGMDGRDVRVAVLHGLPPSPESYVOAV 802
Db 278 AYHAELETSOREQVRAFORNVVAVTIAFGMGINKSNVAFVHFDPJPRJESISTOET 337
QY 803 GRAGDGOFAHEHLELOOGS-----EDURELRHHVHADSTDFLAVKRLV 846
Db 338 GRAGDDJPAEAVLYEPADYAMLKHILKEKESPORQIEALKLQAIIGFAESQTCRRILV 397
QY 847 -----QRVFPACTC-TCTRPPSQEGAVGSE 872
Db 398 LUNYFEGHQKPCQNCIDICLDPKPYQJGIDIAOK 431

```

Search completed: March 19, 2005, 01:12:35
Job time : 169 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2005, 00:59:02 ; Search time 50 Seconds
(without alignments)
2324.597 Million cell updates/sec

Title: US-09-889-325-4

Perfect score: 6424

Sequence: 1 MERLRDVERLRQAMERARRR.....HLSFHALVGLATRELLQVAR 1208

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 79: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	872	13.6	941	2	A86404 probable protein A
2	631.5	9.8	610	1	BVECRQ DNA helicase recQ
3	631.5	9.8	611	2	F86069 ATP-dependent DNA
4	631.5	9.8	611	2	H91222 ATP-dependent DNA
5	609.5	9.5	609	2	AH0917 ATP-dependent DNA
6	593	9.2	610	2	A10466 ATP-dependent DNA
7	581.5	9.1	620	2	E82351 ATP-dependent DNA
8	560.5	8.7	602	2	H97365 DNA helicase Xf138
9	560.5	8.7	602	2	AH2583 ATP-dependent DNA
10	552	8.6	590	2	AD1419 ATP-dependent DNA
11	551	8.6	718	2	AE1832 ATP-dependent DNA
12	548	8.5	590	2	AE1794 ATP-dependent DNA
13	545	8.5	712	2	E83226 ATP-dependent DNA
14	542	8.4	496	2	A69691 ATP-dependent DNA
15	525.5	8.2	1328	2	S62467 ATP-dependent DNA
16	523	8.1	824	2	G75413 DNA helicase RecQ
17	521	8.1	615	2	G87678 ATP-dependent DNA
18	514	8.0	593	2	B89844 hypothetical prote
19	513	8.0	592	2	H86850 ATP-dependent DNA
20	510.5	7.9	676	2	T34609 probable helicase
21	507	7.9	480	2	AB2411 ATP-dependent DNA
22	506.5	7.9	809	2	T20430 hypothetical prote
23	506	7.9	1331	2	T24415 ATP-dependent prote
24	505.5	7.9	478	2	S77358 ATP-dependent DNA
25	505.5	7.9	591	2	F69901 DNA helicase recQ
26	504	7.8	766	2	H81794 ATP-dependent DNA
27	499	7.8	491	2	G83850 ATP-dependent DNA
28	499	7.8	766	2	G81216 ATP-dependent DNA
29	496.5	7.7	1436	2	T14895 DNA helicase 1 - A

30	493.5	7.7	645	2	A82689 DNA helicase Xf138
31	489	7.6	1447	2	S50918 DNA helicase TP51
32	484	7.5	1401	2	T30247 Werner syndrome pr
33	479.5	7.5	1401	2	T17452 Werner syndrome pr
34	478.5	7.4	714	2	G97230 DNA helicase homol
35	476	7.4	665	2	B86243 DNA helicase RECQL
36	464.5	7.2	659	1	A55311 DNA helicase RECQL
37	463.5	7.2	1417	2	A57570 Bloom's syndrome r
38	458.5	7.1	459	2	B89927 hypothetical prote
39	443	6.9	1031	2	G96634 probable DNA helic
40	437	6.8	467	2	AF1317 ATP-dependent DNA
41	434.5	6.8	467	2	AF1689 ATP-dependent DNA
42	432	6.7	886	2	T16536 hypothetical prote
43	422	6.6	483	2	T51906 related to recQ ge
44	406.5	6.3	607	1	C71367 probable ATP-depen
45	386	6.0	1273	2	E72611 probable ATP-depen

ALIGNMENTS

RESULT 1

A86404 probable protein ATP-dependent DNA helicase RecQ [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #ext_change 31-Mar-2001

C:Accession: A86404
R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkin, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lurie, J.S.; Malt, R.; Marshall, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; PMID:21016719; PMID:11130712

A:Accession: A86404

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-941 <STO>

A:Cross-references: GB:AE005172; NID:g10998928; PID:AAG26068.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 13.6%; Score 872; DB 2; Length 941;
Best:Local Similarity 26.2%; Pred.No. 1.4e-38;

Matches 305; Conservative 165; Mismatches 390; Indels 306; Gaps 43;

QY	93	TPGRSRQGVPIYCGRLKANKLTLOAGPALRRPWPGLRASSKASTPRPGTGPVPSFA	152
DB	15	TPPRD---SFPSSPQLOSFAKRV---PPVSRKM---TSSSSRKRPAP-THPPNPS	62
QY	153	KVYSDPPQLPPOPGRGLQHLQASLSQRLSLDPGMLQRCHSEVPDGLGAPKARPD	212
DB	63	QCAPVPSPYPPPPPPSP---LFTNLPRFICQSDA---RSSSVSSP---SRUCSRAS	111
QY	213	GSEESQLTPGESAVLPGPAGSQGPESAFQEVSRVGSPOPSGCEKRWNEBWESE	272
DB	112	FTSVETKSDGVDFV-----PEPPLVEVIA-----	136
QY	273	AOVQDSGAGPPSBGAGVAVEEDPPG---BPVQAPQPQCPSPSPRYHGLSPSQAR	329
DB	137	-----PPK-----SVRRKPPNLTITLITSPKPVFRSN-----	166
QY	330	AKKAEGTAPLHIFPRLARHDRGVYVYLNKKQGVYVGRALRSRL-----LRQAWKQK	383
DB	167	-GNFVKLNKGR---RGKFPSPKVKGVSKRSRYSRGR	206
QY	384	RKKGCFGGGATVTTKESCFINEQFDHMAACCPRASEEDTDAVGPEPLVSPQCPVPEV	443
DB	207	YKKKADGDSIL---EESDLOKQI-----EDENANGF-----I	238

```

QY 444 PSLDPTVLPLYSIGSPGSLAETPAEYFQALBQLGHOAPRPGOERAVMRLISGISTLLVP 503
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 SSVEDAILAVKTEADENITKLNLVY-----GIDSPDQLOAIKMLIGSSITMLVP 292
QY 504 TGAGKSLCYOLPALLYSRRSPCLTLVVSPLSLMDQVSGLPPLCKACIHSGMTKKORE 563
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 TGAGKSLCYQIDPAML-----PGITLVSPVLSIMDLGHLPSIIGKGLSSQREBEAT 348
QY 564 STLQKIRAAQVAVMLTPREALVAGGLPPAAQLPPAFACIDEANCLSQ-----612
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 349 ETLRLKKEGIIIVLFFSPERLLNVEFLSNFRMSLSVLVAVBAHCVSICVCLSPCLIS 408
QY 613 W-----SHNRPFCYLRV--CKVLBERGVCHFLCLTATATRRATSDVAOHIAVAEEDP 663
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 409 WLLRLAIRSNHFRPSRYMMLKASMLFSEIKACILAMTATITMTLOAVMSLEI-PSIN 467
QY 664 LHGPAPVPTNLHLVSM--DRDTPQALLTLQGRKPPMDISIIYCNREDTERIALLL 720
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 468 LIQKSQLRDNFELSVLSGANRMD--LILMESPPYKEIRSIIVYCKFYQYETDMISKYL 525
QY 721 RCLLAAMVPGSGGAPKTTAAYAHGMSRRRRVQRAFMQGLRVVAVATAFGMGLDR 780
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 526 RD-----NNINAKGYHSGLPKAKORVRIQESFCSNKIKRVVAVATAFGMGLDK 571
QY 781 PDVRAVILHGLPPSPESYVQVAGRAGRDGQPAHCHFLQPGEDLRELRRHYADSTDFL 840
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 572 GIVGAVIHHSVSGSMEEYQEIIGRAGDRSLSYCHLFY--DNDTYIKLMSLAHSDVDY 629
QY 841 AVKRLVQVRPACTCTTRPSEBQAVGGERPVPRQEAELQSHQAAPGRVRCWH 900
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 630 AVGKFLTHVF-----STETKQHE-----KIC--650
QY 901 EHALPTQLTVQALDMPERAIETLLCYLEHPPHMLELATTYTHCCLNPGGPAQLQALA 960
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 651 --SLVIESASQKFDMEKEVMQITLLTLEGEVQYLMRLQNLNICCTLN-----F 697
QY 961 HR---CPPLAVCLAQOLPEDPGGSSSVDFDWKVLVDNSG---WELASVRALCQLQMDH 1014
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 698 HTSISDAQNIIVMLYVQSPPTLARSALIVAILKSHVKGGLVFDIPAYASISCVAT--- 754
QY 1015 EPRTVGRGTGVLVESSELAFLRSPGDLTAEBKD-----QICDFLYGRVQAR 1062
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 755 -----TDVLAELIQL-----KGVTYELKDSAFCTILKSPKELSSHLTKWL 799
QY 1063 ERQALAIRLR--TFQASHVAFPSCGPCLEQODEERSTRKDLGRFEEBEGEPGGMDD 1121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 800 TEIESCKVRKLDIMSSAAVAISVNSTSELSSGAKQTR--SLQSRIFDYFNG-----D 850
QY 1122 ACGPERGQARLDWEDQ---VRCDIROFLSLRPEKFSRAVARIFHGIGSPCYAQVY 1177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 851 EKCDSSSKA-----TQNCAPLRADIKVILQSNRQAKFTPRALARIMHGVSGAPFNSW 904
QY 1178 GQDRRFRWRYLHLSPHALVGLATEEL 1203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 905 SK-THFWGRYVMDVFRVIMEAAQTEL 929
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 2
BVECRQ
DNA helicase recQ - Escherichia coli (strain K-12)
N:Alternate names: DNA-dependent ATPase recQ (EC 3.6.1.-)
C:Species: Escherichia coli
C:Date: 31-Dec-1989 #sequence revision 10-Oct-1997 #text_change 01-Mar-2002
C:Accession: G65186; J50137; A35776; S30712
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G65186
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

```

```

A:Residues: 1-610 <BLAT>
A:Cross-references: GB:AEO00458; GB:U00096; NID:G2367299; PIDN:AACT6825.1; PID:G2367301;
A:Experimental source: strain K-12, substrain M31655
R:Ritino, N.; Nakayama, K.; Nakayama, H.
Mol. Gen. Genet. 205, 298-304, 1986
A:Title: The recQ gene of Escherichia coli K-12: primary structure and evidence for SOS
A:Reference number: J50137; MUID:87115164; PMID:3027506
A:Accession: J50137
A:Molecule type: DNA
A:Residues: 1-256, 'A', 258-610 <IRI>
A:Experimental source: strain K12
A:Note: It is uncertain whether residue 1, 3, 8 (all coded by GTG), or 9 (coded by TTG)
R:Umez, K.; Nakayama, K.; Nakayama, H.
Proc. Natl. Acad. Sci. U.S.A. 87, 5363-5367, 1990
A:Title: Escherichia coli recQ protein is a DNA helicase.
A:Reference number: A35776; MUID:90319113; PMID:2164680
A:Accession: A35776
A:Status: preliminary
A:Molecule type: protein
A:Residues: 4-8 <UMS>
R:Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.
Science 257, 771-778, 1992
A:Title: Analysis of the Escherichia coli genome: DNA sequence of the region from 84.5 t
A:Reference number: S30660; MUID:92358234; PMID:1379743
A:Accession: S30712
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'V', 2-256, 'A', 258-610 <DAN>
A:Cross-references: EMBL:M87049
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1992
C:Genes:
A:Gene: recQ
A:Map position: 85 min
A:Start codon: GTG
C:Function:
A:Description: involved in the recF recombination pathway; its gene expression is under
A:Pathway: recF recombination
C:Superfamily: recQ protein; recQ helicase homology
C:Keywords: ATP; DNA binding; hydrolyase; nucleotide binding; P-loop; recF recombination
F:49-56/Region: nucleotide-binding motif A (P-loop)
F:144-149/Region: nucleotide-binding motif B
F:148-151/Region: DEAH motif
F:369-402/Domain: recQ helicase homology <RHH>

Query Match
Best Local Similarity 37.6%; Score 631.5; DB 1; Length 610;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

QY 474 BQLGHOAPRPGOERAVMRLISGISTLLVLP TGAGKSLCYOLPALLYSRRSPCLTLVVSPL 533
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 21 ETFGVQOFRPGQEBIIDVLSGRDCLVMPVTPGGSLSLYQIDALLNG-----LTVVVSPL 76
QY 534 ISLMDQVSGLEPP-CLKAACIHSGMTKQRESVLDKIRAAQVHVMLTPEALVAGGLPP 592
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 77 ISLMDQVQLOANGAAACLNSTQREQOLEVMVGCRTGQIRLYIAERLMLDNFLSH 136
QY 593 AAQLPVPAPACIDBAHCAISQWGNHFRPCYLRCKVLRERMGVCHFCGLTATRRASDV 652
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 137 LAHNPVLLA-VDEAHCTISQWGHDRPREYALGQ-LRQRFPLPMLATATADDTTRQDI 194
QY 653 ACHLVAEBPDLHGPAPVPTNLHLSVMDRTDQALLTLQKRFQNDST-----703
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 195 VALLG-----LNDPL-----IQIS-SFDRPVRIRYMLM-----EKRFPLDQLRKYVQEQRG 238
QY 704 ----IYCNREDTERIALLRCLLAAMVPGSGGAPKTTAAYAHGMSRRRRVQRAF 760
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 KSGIIVYCNRAVEYDTAAL-----OSKGISAAAYHAGLENNVRADQEKF 284
QY 764 MGOGLRVVAVATAFGMGLDRPDVRAVILHGLPPSPESYVQVAGRAGRDGQPAHCHFL-- 818
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 285 QRDDLIIVAVATAFGMGLKPNRVVAFHDIPRNIESTIYQEGRAGRDGLPDEAMLFYDP 344
QY 819 -----QPGEDLRELRRH--VHADSTDFLAVKRLV-----QRFVPACTC- 855

```

Db 345 ADMAMLRCLREKPOGQ-LQDIERHKLNMGAFAPAEQOTCRRLVLTNYFEGEQBOFGNCD 403
 QY 856 TCTRPPEQEGAVGE 871
 Db 404 ICLDPPKQYDGSSTDAQ 419

RESULT 3

F86069

ATP-dependent DNA helicase [imported] - Escherichia coli (strain O157:H7, substrain EDL95)

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #revision 16-Feb-2001 #text_change 09-Jul-2004

C:Accession: F86069

R:Petra, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew

Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: AB5480; MUID:21074935; PMID:1120551

A:Accession: F86069

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-611 <STO>

A:Cross-References: UNIPROT:O8X8N1; UNIPROT:O8FBM6; GB:AB005174; NID:g12518696; PIDN:AAC

A:Experimental source: strain O157:H7, substrain EDL953

C:Genetics:

A:Gene: recQ

C:Superfamily: recQ protein; recQ helicase homology

Query Match 9.8%; Score 631.5; DB 2; Length 611;
 Best Local Similarity 37.6%; Pred. No. 4,1e-26;
 Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

QY 474 EOLGHQAPRPGQERAVMRLISGISTLLVLTPTGAGKSLCYQLPALYSRRSPCLTLVVSPL 533
 Db 21 EFTFGYQFRPQGEIITDVLSGRDCLVMPGTGSKSLCYQIPALLNG---LTVVVSPL 76
 QY 534 ISLMDQVSGLEPP-CLKAACIHSGMTKQRESVLOKIRAAQVHVMLTPEALVAGAGLEPP 592
 Db 77 ISLMDQVDOQLANGVAAACLNSTQTRQGLEVMTCRGQIRLLYIAPERLMDNLFELH 136
 QY 593 AAQLPPVAFACIDEAHCLISQMSHNPFCYLRCKVLREMGVHCFGLTATATRTASDV 652
 Db 137 LAHNPVLLA-VDEAHCLISQMGHDFRPEYALGO-LRQRPPTLPFALATATDTRTODI 194
 QY 653 AQHLVAEPELDHGPAPVTNLTLSVMDRDTQALLTLQGRFQNDISI----- 703
 Db 195 VRLLG-----LNDPL-----IQIS-SFDRPPIRIVLM-----EKFRPDLQLMRYQEQRG 238
 QY 704 ---IYCNREDTERIALLRCTLHAAMVPGSGGAPKTTAAEYHAGMCSRRRRVQRAF 760
 Db 239 KSGIITCNSRAKVEDTAARL-----QSKGISAAAYHAGLENNVRADVQERF 284
 QY 761 MGOGLRVVATVAFMGKLDPRVAVLHGLPSPFSYVQAVGRAGROGPAHCHLFL-- 818
 Db 285 QRDDLIQIVATVAFMGKINPVRVHFDIPRNISYVQETGRAGRDGLPFAAMLFDYP 344
 QY 819 -----OPQGEDLRELRRH---VHADSTDFLAIVKLV-----QRVFPAATC- 855
 Db 345 ADMAMLRCLREKPOGQ-LQDIERHKLNMGAFAPAEQOTCRRLVLTNYFEGEQBOFGNCD 403
 QY 856 TCTRPPEQEGAVGE 871
 Db 404 ICLDPPKQYDGSSTDAQ 419

RESULT 4

H91222

ATP-dependent DNA helicase [imported] - Escherichia coli (strain O157:H7, substrain R1MD

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004

C:Accession: H91222

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: H91222

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-611 <HAY>

A:Cross-References: UNIPROT:O8X8N1; UNIPROT:O8FBM6; GB:BA000007; PIDN:BA038175.1; PID:g1

A:Experimental source: strain O157:H7, substrain R1MD 0509952

C:Genetics:

A:Gene: ECG4752

C:Superfamily: recQ protein; recQ helicase homology

Query Match 9.8%; Score 631.5; DB 2; Length 611;
 Best Local Similarity 37.6%; Pred. No. 4,1e-26;
 Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

QY 474 EOLGHQAPRPGQERAVMRLISGISTLLVLTPTGAGKSLCYQLPALYSRRSPCLTLVVSPL 533
 Db 21 EFTFGYQFRPQGEIITDVLSGRDCLVMPGTGSKSLCYQIPALLNG---LTVVVSPL 76
 QY 534 ISLMDQVSGLEPP-CLKAACIHSGMTKQRESVLOKIRAAQVHVMLTPEALVAGAGLEPP 592
 Db 77 ISLMDQVDOQLANGVAAACLNSTQTRQGLEVMTCRGQIRLLYIAPERLMDNLFELH 136
 QY 593 AAQLPPVAFACIDEAHCLISQMSHNPFCYLRCKVLREMGVHCFGLTATATRTASDV 652
 Db 137 LAHNPVLLA-VDEAHCLISQMGHDFRPEYALGO-LRQRPPTLPFALATATDTRTODI 194
 QY 653 AQHLVAEPELDHGPAPVTNLTLSVMDRDTQALLTLQGRFQNDISI----- 703
 Db 195 VRLLG-----LNDPL-----IQIS-SFDRPPIRIVLM-----EKFRPDLQLMRYQEQRG 238
 QY 704 ---IYCNREDTERIALLRCTLHAAMVPGSGGAPKTTAAEYHAGMCSRRRRVQRAF 760
 Db 239 KSGIITCNSRAKVEDTAARL-----QSKGISAAAYHAGLENNVRADVQERF 284
 QY 761 MGOGLRVVATVAFMGKLDPRVAVLHGLPSPFSYVQAVGRAGROGPAHCHLFL-- 818
 Db 285 QRDDLIQIVATVAFMGKINPVRVHFDIPRNISYVQETGRAGRDGLPFAAMLFDYP 344
 QY 819 -----OPQGEDLRELRRH---VHADSTDFLAIVKLV-----QRVFPAATC- 855
 Db 345 ADMAMLRCLREKPOGQ-LQDIERHKLNMGAFAPAEQOTCRRLVLTNYFEGEQBOFGNCD 403
 QY 856 TCTRPPEQEGAVGE 871
 Db 404 ICLDPPKQYDGSSTDAQ 419

RESULT 5

AH0917

ATP-dependent DNA helicase (EC 3.6.1.-) [imported] - Salmonella enterica subsp. enterica

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AH0917

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moulie, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.,

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AH0917

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-609 <PAR>

A:Cross-References: GB:AL513382; PIDN:CAD07934.1; PID:g16504479; GSPDB:GN00176

C:Genetics:

A:Gene: recQ

C:Superfamily: recQ protein; recQ helicase homology

R.Glaeser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Science 294, 849-852, 2001

A:Authors: Kretz, J.; Kuhn, M.; Kunat, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria species*.
A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AD1419
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-590 <GLA>
A:Cross-references: UNIPROT:O8Y354; GB:NC_003210; PIDN:CAD00970.1; PID:g16412257; GSPDB:A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2757
C:Superfamily: recQ protein; recQ helicase homology

Query Match 8.6%; Score 552; DB 2; Length 590;
Best Local Similarity 36.4%; Pred. No. 6, 2e-22;
Matches 132; Conservative 66; Mismatches 125; Indels 40; Gaps 9;

QY 474 EQLGQAAPPGGERRAVMRILSGISTLLVPTGAGKSLCYQLPALLYSRRSPCLTLVSP 533
DB 11 QNFGYQDFRNGQVDVYSKLCAGEBDTLAIPTGGKSLCYQIPALFDG---LTIIVSPL 66
QY 534 ISLMDQVSGL-PPCLKAACHSGMTRKQRESVLQKIRAAQVHVMLEPVALVAGGLPP 592
DB 67 ISLMDQVVALVSEGIATFTINSTLTNREIDRLDAASGELKMLYIAERLETPEFQRL 126
QY 593 AAOLPPVAFACIDEAHCLSQWSHNPFCYLAVCYLREMGVHCFGLTATATRTASDV 652
DB 127 IQQVPLSLFA-IDEAHCISQMGHDFRPSYLTLCDSLDKMTRRPLVALTATATQAVSDI 185
QY 653 AQLHVAEERPDHGAAPVT-----NLHLSVSMRPTDQALLTLQKGFQVL-DSITTY 706
DB 186 CELLTIRAD-----SVKTFGRSDNLAPQVYKQDKXYLLDYLT---KNTSEGITTY 235
QY 707 CNRREDTERIALLRCTLHAAMVPGSGRAPPTTAA--YHAGMCSRRERRVQRAFMQQL 764
DB 236 ASTRKEVERLHSLFQ-----KKGVESGMTHGQMDIARKMQEKLYVD 279
QY 765 LRVVATVAFMGGLDRPDVRAVLHLGLPPSPFSYVQAVGRAGRDQPAHCHLFLQPGSD 824
DB 280 IRVIVATVAFMGINKSNVRFVTHVNIPTNIEAVYQEAARAGRDGVPSCILLFSPQDSR 339
QY 825 LRE 827
DB 340 IQQ 342

RESULT 11
AE1832
ATP-dependent DNA helicase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AE1832
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Ana*
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE1832
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-718 <KUR>
A:Cross-references: UNIPROT:Q82093; GB:BA000019; PIDN:BAW77729.1; PID:g17135183; GSPDB:C
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0205
C:Superfamily: recQ protein; recQ helicase homology

Query Match 8.6%; Score 551; DB 2; Length 718;
Best Local Similarity 35.6%; Pred. No. 8, 7e-22;
Matches 153; Conservative 71; Mismatches 148; Indels 58; Gaps 17;

QY 471 QALF-QLGHQAAPPGGERRAVMRILSGISTLLVPTGAGKSLCYQLPALLYSRRSPCLTLV 529
DB 9 QALKTHFYQDFRNGQVDVYSKLCAGEBDTLAIPTGGKSLCYQIPALFDG---KQGLTYV 64
QY 530 VSPSLTMDQVSGL-PPCLKAACHSGMTRKQRESVLQKIRAAQVHVMLEPVALVAGG 588
DB 65 VSPSLTMDQVVALVSEGIATFTINSTLTNREIDRLDAASGELKMLYIAERLETPEFQRL 124
QY 589 GLP---PAOLPPVAFACIDEAHCLSQWSHNPFCYLAVCYLREMGVHCFGLTATAT 645
DB 125 FLPLDLVKEKVGISIPALDEAHCVSEWGHDFRPEY-RQLSLRGRYNNVPLATATAT 183
QY 646 RRTASDVQHLVAEERPDHGAAPVTTHLSVSMRPTDQALLTLQKGFQVL-DSITTY 705
DB 184 DNRADITIQQLGL-KQPSIHLSFNRQMLYYEVR--PKSKQAVALLBLIR-DNEGSTII 239
QY 706 YCNREDTERIALLRCTLHAAMVPGSGRAPPTTAAVYHAGMCSRRERRVQRAFMQQL 765
DB 240 YCLTKKVBELTFKL-----QDKISALSYHAGLPDESKQVTRIRQDV 285
QY 766 RRVVATVAFMGGLDRPDVRAVLHLGLPPSPFSYVQAVGRAGRDQPAHCHLFL----- 818
DB 286 RVMVATVAFMGINKSNVRFVTHVNIPTNLSYQESGRAGRDGSAASCTIFFSPGDIKT 345
QY 819 -----QPGEDL--RELRRHV-HADSTDFLAVKLVQ-----RFPFA-C-TC-TCT 858
DB 346 IEWSIEQKTDPOEOLIAQQLRMQMDVAGTD--CRRTIQGLYGRGFRPFGCNCNCR 402
QY 859 RPPSEQEGAV 868
DB 403 YPKWQDMWTI 412

RESULT 12
AE1794
ATP-dependent DNA helicase homolog lin2900 [imported] - *Listeria innocua* (strain Clp1)
C:Species: *Listeria innocua*
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AE1794
R:Glaeser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Kretz, J.; Kuhn, M.; Kunat, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria species*.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AE1794
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-590 <GLA>
A:Cross-references: UNIPROT:O92620; GB:AL592022; PIDN:CAC98125.1; PID:g16415441; GSPDB:C
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: lin2900
C:Superfamily: recQ protein; recQ helicase homology

Query Match 8.5%; Score 548; DB 2; Length 590;
Best Local Similarity 36.5%; Pred. No. 1e-21;
Matches 132; Conservative 65; Mismatches 127; Indels 38; Gaps 9;

QY 474 EQLGQAAPPGGERRAVMRILSGISTLLVPTGAGKSLCYQLPALLYSRRSPCLTLVSP 533
DB 11 QNFGYQDFRNGQVDVYSKLCAGEBDTLAIPTGGKSLCYQIPALFDG---LTIIVSPL 66
QY 534 ISLMDQVSGL-PPCLKAACHSGMTRKQRESVLQKIRAAQVHVMLEPVALVAGGLPP 592
DB 67 ISLMDQVVALVSEGIATFTINSTLTNREIDRLDAASGELKMLYIAERLETPEFQRL 126

QY 593 AAOLPPVAFACIDEAHCLSMWSHNPAPCYLRYCKVLREMGVHCFGLTATATRTASPDY 652
 DB 127 IQGVPISTLFA-IDENACISOMGHDFRPSLISLSDSKTRRLVATLITATQAVSDI 185
 QY 653 AOHLLAAVEEPDLHGPAVPPT-----NLHLSVSMRDRDQALTLTLOGKRFONLDSIIIVC 707
 DB 186 CRLTKI--KPN-----SVIKTGFERDYLAFQVAVGQDKKYLIDYLRK-----NSTESGIIVA 236
 QY 708 NRREDTERIALLRITCLHAAWPGSGGAPKTTAE--YHAGMCSERRRVQAPFQOGL 765
 DB 237 STRKEERHISFL-----KKVESEGWHGWTDTARKDWEKFLYDDI 280
 QY 766 RVVAVTAVAGMGDLRDPVAVLHGLPSPFESYVQAVGRDGPAPHCHLFLQPGEDL 825
 DB 281 RIVATNAFGMGINKSNVAFVHYHNPRIEATVQAGRAGRDGVSDDLFLFSDSKI 340
 QY 826 RE 827
 DB 341 QQ 342

RESULT 13

ES3226
 ATP-dependent DNA helicase RecQ PA3344 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
 C:Accession: E83226
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lam,
 J.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 gen
 A:Reference number: A82950; PMID:20437337; PMID:10984043
 A:Accession: E83226
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-712 <STO>
 A:Cross-references: UNIPROT:Q9HY01; GB:AE004757; GB:AE004091; NID:G9949476; PIDD:ANG0673
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: recQ; PA3344
 C:Superfamily: recQ protein; recQ helicase homology

Query Match 8.5%; Score 545; DB 2; Length 712;
 Best Local Similarity 28.5%; Pred. No. 1.8e-21;
 Matches 221; Conservative 102; Mismatches 294; Indels 158; Gaps 32;
 QY 471 QALBOL-----GHOAFRPGQERAVMRILSGISTLVLPTGAGSKLCYQLPALYSRRSPCL 526
 DB 4 QALRLKGVFGYDARFNGQARIIEVAEGDGLVIMPTGGKSLCFQVALLAREG----L 59
 QY 527 TLVSPPLSLMDQVS-----GLPCLKACIHSGMTKRRKRSVLOKIRAAOVHVLMLTP 581
 DB 60 TVVVSPLTALMEDQVATDELGV-----AVALNSTLNPQQRDIABRLGRGIRKILLYLAP 115
 QY 582 EALVAGGIPPAACLPVAFACIDEAHCLSMWSHNPAPCYLRYCKVLREMGVHCFGLT 641
 DB 116 ERLVPRMLATLQRLPVGIFA-IDENACISOMGHDFRPSLISLSDSKTRRLVATLITATQAVSDI 173
 QY 642 ATATRTASDVAVQHLLAAVEEPDLHGPAVPPTMHLVSVMRDRDQALTLTLOGKRFONL 701
 DB 174 ATADKRTHEMTQRHLQNAEQFLSFDPRNIFHYIVREGRKQ-LIGFLSERG--D 229
 QY 702 SIITICNRREDTERIALLRITCLHAAWPGSGGAPKTTAEYHAGMCSERRRVQAPFQ 761
 DB 230 AGIVVCLSRKVEAEAEFL-----GNQG-----FPALPYHAGLSNELRAHQKREPL 275
 QY 762 OGQLVWVAVTAVAGMGDLRDPVAVLHGLPSPFESYVQAVGRDGPAPHCHLFLQPG 821
 DB 276 NEEGLIMATLAFNGSIDKPNRPFVANHDLPLSLELYVQETRAGDGLPADAMWAVGLQ 335
 QY 822 GEDLLERRHVADSTDLAVKRLVQVRPACTCTCTRPPEOEGAVG---GERVP--- 875

DB 336 ---DVILLRQMQSSSD-BRHKRVENHKLAMALCEETRCRQALLAYFDEEMPOPGH 392
 QY 876 -----KYPPOEABOLSHQAP---GPRRVCKGERALPIQTVOALDPEBAITLLC 925
 DB 393 CINCVDGVEWTWATESARQALSAIYRSGQRYGVGH-----LVDTLLGRETEKIRSL-- 443
 QY 926 YLEHPR-----HW-----LELLATYTHCRINCPCGGPQALAHRCPL-- 966
 DB 444 ---GHQHLAVFGIGKRGDEWRTLFRQLVAGLADVDLDGFRG---LRLTEACRPLR 496
 QY 967 ---AVCLAQQLBEDPQSSS---SVEFDMLVDSMGWE-LASVRAALCOLQMDPEPT 1018
 DB 497 GEVRLERLDLPRQAKSSSGGASAGSLVSEBEMMALRLKLAE---EHSVPP 553
 QY 1019 GVRRTGVLVPESELAFLHRSFGDLTAEEKDQICPLYGRVQAREQALRLRTFOAF- 1077
 DB 554 YVIFPDATLEW--LRSPRSLSDMAQ-----VSGVGR-----KLERYQALFL 595
 QY 1078 -----HSVAFPGC-GPCLGQDEERSTRKLDLGRYFEEBEGQEPGG 1118
 DB 596 DVLTSPPAAPAPPODLRLHLSLACAGTPAQIARQLMCKSEKNVAMLAIAQGVSL 655
 QY 1119 MEDAQPPEPGQARLDQMDQVRCDIROFL---SLRP---EEXTSSRAVARIFH 1165
 DB 656 EQALDLPBELLGIDQ-----AFLEEDGELPVAALBERGKRVPSGLH 700

RESULT 14

A69691
 ATP-dependent DNA helicase homolog - Bacillus subtilis
 N:Alternate names: ATP-dependent DNA helicase recQ
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
 C:Accession: A69691
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Beret
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Evington, J.; Fabre, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.
 lech, J.; Harwood, C.R.; Knecht, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois,
 A.; Lauter, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maesuda, S.; Maueel
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
 A.; Schaefer, S.; Schroeder, R.; Scrofano, F.; Sekiguchi, J.; Sekowska, A.; Seto
 akuch, M.; Tanakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipac, A.; Yamamoto, K.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; PMID:98044033; PMID:9384377
 A:Accession: A69691
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-496 <KUN>
 A:Cross-references: UNIPROT:P50729; GB:Z99115; GB:Z99116; GB:AL009126; NID:G2634723; PID
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: recQ
 C:Keywords: ATP; nucleotide binding; P-loop
 F:38-45/Region: nucleotide-binding motif A (P-loop)
 F:132-137/Region: nucleotide-binding motif B
 F:136-139/Region: DNA motif

Query Match 8.4%; Score 542; DB 2; Length 496;
 Best Local Similarity 36.8%; Pred. No. 1.7e-21;
 Matches 137; Conservative 68; Mismatches 117; Indels 50; Gaps 12;
 QY 471 QALBOL-----LGHQAFRPGQERAVMRILSGISTLVLPTGAGSKLCYQLPALYSRRSPCLTY 529
 DB 6 QTLVQFFFTSKKQDQIISILSGKDTIALPLTGGKSLCYQLPGYMLDG---MWLI 61
 QY 530 VSPPLSLMDQVSGLPPC--LKAACIHSGMTKRRKRSVLOKIRAAOVHVLMLTPPALVGA 587

Db 62 VSPILSLMEDVOQULKARGEKRAALNSMLNROEQFVLEHIIH--RYKFLVLSPEALOSP 119
 QY 588 GGLPPAAQLPPVPAFCIDIEAHCLISQWNSHNPFCYLRVCKVLREBMGVHCFGLTATATRR 647
 Db 120 YLEELTKSV-PLSLFVIDEABHCTISEMGHDFRDPYKLCQ-LRKKGHPVPLTLTATATKE 177
 QY 648 TASDV-----AQHLVAVEEPDLHGPAVPVPTNHLVSMSDRDTPDQALLTLQGRFQ 698
 Db 178 TLQDQWNLLELGHAVRHLSNVNR-----NIALRVENAADTAEKIDRVIQ--LVE 225
 QY 699 NLDSI-IYCNRRRETERIAALLRTCLHAANVPGGGAPKTTAA--YHAGMCRERRR 755
 Db 226 NLQGGIYVCPTRKMKELAGEIKS-----KTSRADPFHGGLESQDRIL 270
 QY 756 VQARMQGLRVVAVTVAFCMGDLDPDYRAVLHLGLPSPFESYQAVAGRAGRDQPAHCH 815
 Db 271 IQQDFIHQDLVICTTNAFGMGVDKPDIRYVHFFHPLPOTAEKFMQEIQRAGDKRPSVI 330
 QY 816 LFLQPGESDLRE 827
 Db 331 LLRAVGDFELQE 342

RESULT 15

S62467

ATP-dependent DNA helicase hus2 - fission yeast (Schizosaccharomyces pombe)

C.Species: Schizosaccharomyces pombe

C.Date: 16-May-1996 #sequence revision 13-Mar-1997 #text_change 09-Jul-2004

C.Accession: S62467; T38578; T43265

R.Badcock, K.; Churcher, C.M.

submitted to the EMBL Data Library, October 1995

A.Reference number: S62445

A.Accession: S62467

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-1328 <BAD>

A.Cross-references: UNIPROT:Q09811; EMBL:254354; NID:g1019398; PIDN:CAA91177.1; PID:g101

R.Badcock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, October 1995

A.Reference number: Z21745

A.Accession: T38578

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-1328 <BA2>

A.Cross-references: EMBL:254354; PIDN:CAA91177.1; GSPDB:GN00066; SPDB:SPAC2G11.12

A.Experimental source: strain 972h; cosmid c2G11

R.Stewart, E.; Chapman, C.R.; Al-Khodairy, F.; Carr, A.M.; Enoch, T.

EMBO J. 16, 2682-2692, 1997

A.Title: Rqh1+, a fission yeast gene related to the Bloom's and Werner's syndrome genes,

A.Reference number: Z07997; MUID:97327548; PMID:9184215

A.Accession: T43265

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-1328 <STB>

A.Cross-references: EMBL:Y09426; NID:g1684753; PIDN:CAA70577.1; PID:g1684754

C.Genetics:

A:Gene: hus2; SPAC2G11.12

A:Map position: 1L

C:Superfamily: Bloom's syndrome helicase; recQ helicase homology

C:Keywords: ATP; nucleotide binding; P-loop

F:541-548/Region: nucleotide-binding motif A (P-loop)

F:647-652/Region: nucleotide-binding motif B

F:651-654/Region: DEAH motif

F:875-914/Domain: recQ helicase homology <RHH>

Query Match

Best Local Similarity 37.1%; Score 525.5; DB 2; Length 1328;

Matches 143; Conservative 49; Mismatches 154; Indels 39; Gaps 11;

QY 502 LPTGAGKSLCYQPLALVSRSPCLTLVSPILSLMDQVSGLP---PCLKAACIHSGM 557
 Db 541 MPTGGKSLCTQLPRAVIEGASRGVTLVISPILSLMDQDLRLKRLNIPSLPISGEOPAD 600
 QY 558 TRKQRESVLOKIRAAQVNVMLTPALVAGG---LPPAQLPPVPAACIDEAHCLSQW 613
 Db 601 ERROVSPFLM-AKVNLVTLVTPPEGLASNGATRVLSLYERKLLARIVIDEACHSHW 659
 QY 614 SHNPPCYLRVCKVLREBMGVHCFGLTATATRTASDVAQHLVAVEEPDLHGPAVPPTN 673
 Db 660 GHDFRPDYKQ-LGLRDPYQGIPPMALTRTANETVKDIIWTLRMENCLBLKSSFNRR-N 717
 QY 674 LHLVSMDRDTDQALLTLQGRFPQNLDSIIYCNRRRETERIAALLFTCLH-AAWPGS 732
 Db 718 LFYEIKPKDLYTELRFISNGHLH--ESGIYCLSRFSCQVAAKLNDYGLKAM----- 771
 QY 723 GGRAPKTAELVHAGMCRERRRRQRAHQGLRVVAVTVAFCMGDLDPDYRAVLHLGLP 792
 Db 772 -----HYHAGLEVERQRIQNEWQSGSYKIIVATIAFGMGVDKGDVRFVTHSPF 821
 QY 793 PSFESYQAVGRAGRDQPAHCHLP 817
 Db 822 KSLBGTYQETGRAGRDGKPAHCIMF 846

Search completed: March 19, 2005, 01:09:44

Job time : 54 secs

QY 443 VPSLPTVLELVSLSPGGLAETPAVFPALQOLGH-QAFRPGQERAVNRITLSGISTLLV 501
 Db 491 MPSLUD---PMLSYPMW-----KEVLGCLKHKFFHLKGRFRQQLERAINGTLSGKDVFTL 540